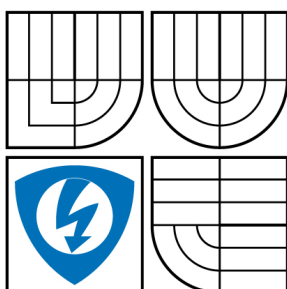




**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ**  
**ÚSTAV TELEKOMUNIKACÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

# NEDOURČENÁ SLEPÁ SEPARACE ZVUKOVÝCH SIGNÁLŮ

UNDERDETERMINED BLIND AUDIO SIGNAL SEPARATION

**DOKTORSKÁ PRÁCE**  
DOCTORAL THESIS

**AUTOR PRÁCE**  
AUTHOR

Ing. JAN ČERMÁK

**VEDOUCÍ PRÁCE**  
SUPERVISOR

prof. Ing. ZDENĚK SMÉKAL, CSc.

BRNO 2008

# LICENČNÍ SMLOUVA

## POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

### 1. Pan/paní

Jméno a příjmení: Ing. Jan Čermák  
Bytem: Budovatelská 1984, 39701, Písek - Vnitřní Město  
Narozen/a (datum a místo): 10.9.1979, Písek

(dále jen "autor")

a

### 2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií  
se sídlem Údolní 244/53, 60200 Brno 2  
jejímž jménem jedná na základě písemného pověření děkanem fakulty:

.....

(dále jen "nabyvatel")

## Článek 1

### Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- ☒ disertační práce
- ☐ diplomová práce
- ☐ bakalářská práce

jiná práce, jejíž druh je specifikován jako .....

(dále jen VŠKP nebo dílo)

Název VŠKP: Nedourčená slepá separace zvukových signálů

Vedoucí/školicitel VŠKP:

Ústav: Ústav telekomunikací

Datum obhajoby VŠKP: .....

VŠKP odevzdal autor nabyvateli v:

- ☒ tištěné formě - počet exemplářů 1
- ☒ elektronické formě - počet exemplářů 1

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.

3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.

4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

## **Článek 2**

### **Udělení licenčního oprávnění**

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
  - ☒ ihned po uzavření této smlouvy
  - ☐ 1 rok po uzavření této smlouvy
  - ☐ 3 roky po uzavření této smlouvy
  - ☐ 5 let po uzavření této smlouvy
  - ☐ 10 let po uzavření této smlouvy(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

## **Článek 3**

### **Závěrečná ustanovení**

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne: .....

.....

Nabyvatel

.....

Autor

## **Abstrakt**

V reálných situacích se často setkáváme s případem, kdy je v neznámém prostředí smícháno více zvukových signálů dohromady. Aby mohly být tyto signály správně interpretovány je nutné užitečné signály extrahovat z jejich směsi. Disciplína zabývající se tímto problémem se nazývá slepá separace signálů.

Tato práce pojednává o vícekanálové slepé separaci zvukových signálů v reálném prostředí, a to i v případě kdy počet zdrojů přesáhne počet senzorů. První částí disertační práce je úvodem do slepé separace signálu. Je zde provedena analýza současného stavu separačních metod. Na základě získaných poznatků jsou navrženy separační systémy implementující více-stavové časově-frekvenční maskování. Tyto metody však stále provádí nelineární změny ve spektru, jež mohou způsobovat hudební šum. Jsou tedy představeny další metody potlačující hudební šum. Tyto metody kombinují časově-frekvenční binární maskování a adaptivní metody směřování přijímací charakteristiky. Výsledkem je nový lineární separační systém aplikovatelný i v případě kdy počet zdrojů převyšuje počet senzorů. V závěrečné části práce jsou navržené separační systémy vyhodnoceny pomocí objektivních i subjektivních testů.

## **Klíčová slova**

Nedourčená slepá separace signálů, směřování přijímací charakteristiky, směrovač, časově-frekvenční maskování, analýza řídkých komponent, řídké signály, hudební šum, zvýrazňování signálu, analýza nezávislých komponent.

## **Abstract**

We often have to face the fact that several signals are mixed together in unknown environment. The signals must be first extracted from the mixture in order to interpret them correctly. This problem is in signal processing society called blind source separation.

This dissertation thesis deals with multi-channel separation of audio signals in real environment, when the source signals outnumber the sensors. An introduction to blind source separation is presented in the first part of the thesis. The present state of separation methods is then analyzed. Based on this knowledge, the separation systems implementing fuzzy time-frequency mask are introduced. However these methods are still introducing nonlinear changes in the signal spectra, which can yield in musical noise. In order to reduce musical noise, novel methods combining time-frequency binary masking and beamforming are introduced. The new separation system performs linear spatial filtering even if the source signals outnumber the sensors. Finally, the separation systems are evaluated by objective and subjective tests in the last part of the thesis.

## **Key words**

Underdetermined blind source separation, beamforming, beamformer, time-frequency masking, sparse component analysis, sparse signal, musical noise, signal enhancement, independent component analysis.

### **Bibliografická citace**

ČERMÁK, J. *Nedourčená slepá separace zvukových signálů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 94 s.  
Vedoucí disertační práce prof. Ing. Zdeněk Smékal, CSc.

### **Prohlášení o původnosti**

Prohlašuji, že svou disertační práci na téma „Nedourčená slepá separace zvukových signálů“ jsem vypracoval samostatně pod vedením vedoucího disertační práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené disertační práce dále prohlašuji, že v souvislosti s vytvořením této disertační práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení §11 a autorského zákona č. 121/2000 Sb., včetně možných trestně právních důsledků vyplývajících z ustanovení §152 trestního zákona č. 140/1961 Sb.

V Brně dne 1. 10. 2008

.....

Ing. Jan Čermák

## **Poděkování**

Děkuji vedoucímu disertační práce prof. Ing. Zdeňku Smékalovi, CSc. za podnětné připomínky a odbornou pomoc při zpracování disertační práce. Dále děkuji Ing. Robertu Víchovi, DrSc. za cenné rady v oblasti zpracování signálů a odborníkům z NTT Communication Science Laboratories, hlavně Shoko Araki, Hiroshi Sawada a Shoji Makino, za poskytnutí simulačního prostředí, odbornou pomoc a vřelý přístup. Děkuji také rodičům a manželce za podporu a trpělivost.



# Obsah

Seznam obrázků .....	11
Seznam tabulek .....	12
1 Úvod .....	13
1.1 Směšovací prostředí .....	15
1.2 Separační proces .....	17
1.2.1 Jednakanálové metody .....	18
1.2.2 Vícekanálové metody .....	18
1.2.3 Rozdělení podle počtu zdrojů a senzorů .....	20
2 Přehled současného stavu problematiky .....	22
2.1 Analýza nezávislých komponent .....	22
2.2 Tvarování přijímací charakteristiky .....	25
2.3 Analýza řídkých komponent v slepé separaci signálů .....	26
2.4 Souhrn .....	31
3 Cíle disertační práce .....	33
4 BSS systémy založené na časově-frekvenčním maskování .....	34
4.1 Systém MENUET .....	34
4.2 Systém MENUET s více stupni příslušnosti .....	37
4.2.1 Systém MENUET s měkkou maskou .....	38
4.2.2 Systém MENUET s fuzzy maskou .....	40
4.3 Experimentální část .....	42
4.4 Souhrn .....	45
5 Kombinace tvarovačů přijímací charakteristiky a systému MENUET .....	46
5.1 Tvarovač přijímací charakteristiky typu Max SNR .....	46
5.2 Tvarovač přijímací charakteristiky typu LCMV .....	48
5.3 Úprava systému MENUET .....	49
5.3.1 Určení korelační matice interferencí a korelační matice zdrojového signálu .....	49
5.3.2 Určení směsných vektorů .....	50
5.4 Sestavení separačního systému .....	51
5.5 Experimentální část .....	54
5.6 Souhrn .....	54
6 Kaskádní spojení systému MENUET a tvarovačů přijímací charakteristiky .....	56
6.1 Převod na určenou BSS .....	56
6.2 Pole tvarovačů přijímací charakteristiky .....	58
6.3 Experimentální část .....	62

6.4	Souhrn.....	63
7	Jednokanálové potlačení interferencí v separovaném signálu .....	64
7.1	Experimentální část.....	64
7.2	Souhrn.....	65
8	Srovnání navržených BSS systémů.....	67
9	Závěr.....	75
	Literatura .....	77
	Seznam použitých symbolů.....	82
	Seznam zkratk .....	86
	Seznam příloh.....	88
	Životopis.....	93

## Seznam obrázků

Obr. 1.1.	Slepá separace signálů ve volném prostředí .....	14
Obr. 1.2.	Model šíření zvuku v místnosti.....	16
Obr. 1.3.	Impulsní odezva s dominantní přímou vlnou .....	16
Obr. 1.4.	Příjem rovinného signálu pomocí sensorového pole.....	19
Obr. 2.1.	Blokové schéma slepé separace pomocí ICA.....	24
Obr. 2.2.	Blokové schéma tvarovače v časově-frekvenční oblasti. ....	25
Obr. 2.3.	Blokové schéma separačního systému složeného z $N$ tvarovačů .....	25
Obr. 2.4.	Časový průběh tří promluv – mužský hlas .....	27
Obr. 2.5.	Časové průběhy modulů tří promluv – mužský hlas .....	28
Obr. 2.6.	Separační maska .....	30
Obr. 2.7.	Histogram normalizovaného vektoru.....	31
Obr. 4.1.	Blokové schéma systému MENUET .....	34
Obr. 4.2.	Maskování pozorovaného signálu systémem MENUET .....	36
Obr. 4.3.	Normalizovaný vektor a histogram separační masky MENUETu .....	37
Obr. 4.4.	Stupeň příslušnosti k shluku $C_k$ .....	39
Obr. 4.5.	Maskování pozorovaného signálu systémem SMMENUET .....	39
Obr. 4.6.	Normalizovaný vektor a separační maska systému SMMENUET.....	40
Obr. 4.7.	Maskování pozorovaného signálu systémem FMMENUET .....	41
Obr. 4.8.	Normalizovaný vektor a separační maska systému FMMENUET.....	42
Obr. 4.9.	Uspořádání experimentálního systému .....	43
Obr. 5.1.	Blokové schéma systému MMENUET.....	51
Obr. 5.2.	Blokové schéma BSS systému kombinujícího tvarovače a systém MMENUET .....	51
Obr. 5.3.	Navržené prostorové filtry BSS systému kombinujícího LCMV tvarovače a systém MMENUET, DC uspořádání.....	52
Obr. 5.4.	Navržené prostorové filtry, obdoba obr. 5.3 pro UDC uspořádání.....	53
Obr. 6.1.	Blokové schéma systému pro nedourčenou BSS.....	57
Obr. 6.2.	Navržené prostorové filtry BSS systému kombinujícího MMENUET a LCMV tvarovače v kaskádním zapojení.....	58
Obr. 6.3.	Pole tvarovačů.....	59
Obr. 6.4.	Blokové schéma BSS systému založeného na BA .....	60
Obr. 6.5.	Navržené prostorové filtry pro $k$ -tý sloupec BA.....	61
Obr. 8.1.	Výsledky subjektivního testu, DC uspořádání, reference MENUET .	70
Obr. 8.2.	Výsledky subjektivního testu, UDC uspořádání reference MENUET	70
Obr. 8.3.	Výsledky subjektivního testu, DC uspořádání, reference BA2 .....	71
Obr. 8.4.	Výsledky subjektivního testu, UDC uspořádání, reference BA2 .....	71
Obr. 8.5.	Výsledky ASR testu. DC uspořádání.....	73
Obr. 8.6.	Výsledky ASR testu. UDC uspořádání.....	73

## Seznam tabulek

Tab. 4.1.	Výsledky experimentů, nelineární systémy založené na MENUETu.	44
Tab. 5.1.	Výsledky experimentů, kaskádní zapojení tvarovačů a MENUETu ..	54
Tab. 6.1.	Výsledky experimentů, převod UDC na DC uspořádání.....	62
Tab. 6.2.	Výsledky experimentů, pole tvarovačů .....	62
Tab. 7.1.	Výsledky experimentu, jednokanálové potlačení interferencí.....	65
Tab. 8.1.	Výsledky experimentu, vyhodnocení metodou PESQ.....	68
Tab. P.1.	Výsledky subjektivního testu, DC usp., reference MENUET .....	89
Tab. P.2.	Výsledky subjektivního testu, UDC usp., reference MENUET .....	89
Tab. P.3.	Výsledky subjektivního testu, UD uspořádání, reference BA2 .....	90
Tab. P.4.	Výsledky subjektivního testu, UDC uspořádání, reference BA2 .....	90
Tab. P.5.	Výsledky ASR testu, DC uspořádání.....	91
Tab. P.6.	Výsledky ASR testu, UDC uspořádání.....	91

# 1 Úvod

Zvukové signály jsou součástí každodenního života. Jsou zaznamenávány sluchovými orgány a předávány dále k zpracování do mozku. Zpracováním se rozumí výběr vhodného signálu, jeho správné porozumění a interpretace. V běžných situacích se však často setkáváme s případy, kdy je současně aktivních více zvukových signálů. Tyto signály jsou smíchány dohromady, což snižuje schopnost porozumění užitečnému signálu. Mnohdy nastává také případ, kdy je více užitečných signálů aktivních v jeden okamžik a posluchač není schopen zaznamenat ani jeden z nich. Takovéto situace jsou běžné např. během telefonických konferencí a společenských akcí.

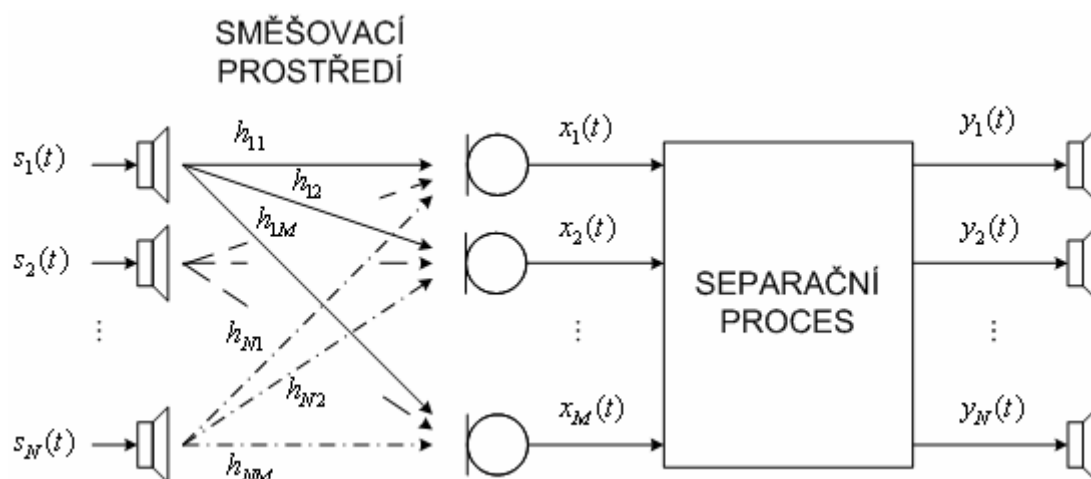
Problém interpretace signálů, jež jsou tvořeny ze směsí užitečných a rušivých signálů není problémem pouze komunikace člověk – člověk, ale také komunikace člověk – počítač. Automatické rozpoznávače řeči jsou velmi závislé na kvalitě přijaté promluvy. Degradace zdrojového signálu interferencemi často vede k snížení úspěšnosti rozpoznávání.

Porozumění přijatému signálu může být zvýšeno, pokud zamezíme vzniku rušivých interferencí a překrytí užitečných signálů. Toto ovšem není v praktických situacích vždy možné. Jinou možností jak zvýšit porozumění přijatého signálu je v sofistikovaných metodách separace signálů.

Slepá separace signálů (BSS) je disciplína zabývající se separací zdrojových signálů z jejich směsí, a to bez jakýchkoli znalostí o směřovacím prostředí. Na obr. 1.1 je uveden příklad slepé separace signálů.  $N$  zdrojových signálů je ve volném prostředí smícháno lineární transformací dohromady. Prostředí je snímáno  $M$  senzory. Zaznamenané směsy signálů jsou zpracovány separačním algoritmem, jehož úkolem je oddělit jednotlivé zdrojové signály a to tak, aby separované signály byly co nejméně zkreslené a zároveň obsahovaly co nejméně interferencí. Informace o směřovacím prostředí či zdrojových signálech nejsou apriori k dispozici.

Tato disertační práce pojednává o slepé separaci zvukových signálů (zkráceně pouze signálů), jakými jsou například řeč či hudba, a to se zaměřením na případ, kdy počet zdrojových signálů  $N$  přesáhne počet senzorů  $M$ . Hlavním přínosem této práce je návrh nového BSS systému použitelného i v případě kdy  $N > M$ , který nedegraduje separovaný signál hudebním šumem, vykazuje nízké zkreslení užitečného signálu a zároveň účinně potlačuje interferující signály.

První kapitola této práce je úvodem do slepé separace signálů. Obsahuje nutné podklady k správné interpretaci dalších kapitol. Zaměřuje se nejdříve na popis směřovacího procesu jak ve volném prostoru, tak i v reálném prostředí. Následuje seznámení s procesem separace včetně stručného rozdělení separačních metod.



Obr. 1.1. Slepá separace signálů ve volném prostředí.

Druhá kapitola se zaměřuje na současné vícekanálové separační algoritmy. Popisuje přístupy založené na lineární prostorové filtraci, jejímiž představiteli jsou analýza nezávislých komponent (ICA) a tvarování přijímací charakteristiky (BF), a metody nelineární prostorové filtrace založené na analýze řídkých komponent (SCA).

Třetí kapitola uvádí hlavní cíl disertační práce. Upozorňuje na nedostatky současných metod a seznamuje s požadavky na separační systém, jež by měl být výstupem této disertační práce.

Čtvrtá kapitola se zabývá časově-frekvenčním maskováním. Představuje separační systém s názvem MENUET, z něhož vycházejí všechny ostatní navržené separační systémy. Kromě představení této standardní metody časově-frekvenčního binárního maskování, jsou v této kapitole prezentovány navržené separační systémy využívající více-stavové masky.

Pátá kapitola pojednává o kombinaci BF s časově-frekvenčním maskováním. Najdeme zde popis dvou adaptivních metod BF, modifikaci systému MENUET a způsob sloučení obou separačních metod do jedné separační metody.

Šestá kapitola představuje zdokonalení separačních systémů z páté kapitoly. Navrhuje kaskádní zapojení modifikované metody MENUET a tvarovačů přijímací charakteristiky (z anglického výrazu beamformer). Seznamuje s polem tvarovačů přijímací charakteristiky.

Sedmá kapitola navrhuje metodu jednokanálového zvýraznění separovaných signálů. Metoda využívá časově-frekvenční masky navržené během procesu separace. Může být použita pro vyšší potlačení interferencí ve výstupních signálech separačních systémů navržených v kapitolách pět a šest.

Osmá kapitola se zabývá porovnáním separačních systémů pomocí subjektivních i objektivních metod.

Na závěr, v deváté kapitole, je uveden souhrn přínosů disertační práce včetně návrhů na další výzkum v oblasti separace signálů.

## 1.1 Směšovací prostředí

Mechanické kmitání částic v pružném prostředí se ve slyšitelném frekvenčním pásmu projevuje jako zvuk. Zvuk se od svého zdroje šíří zvukovými vlnami – postupným podélným vlněním. Body pole, v nichž částice vykazují stejnou fázi, nazýváme vlnoplochami.

V homogenním prostředí se kolem bodového nebo kulového zdroje zvuku tvoří kulové vlnoplochy. Ve velké vzdálenosti od zdroje lze v omezeném prostoru, vzhledem k rozměru senzoru, považovat tyto vlnoplochy již za rovinné. Akustické pole můžeme tedy rozlišit na kulové (near-field model) a rovinné (far-field model).

Pro popis směšovacího prostředí uvažujeme volný prostor s rovinným akustickým polem viz obr. 1.1. Ve volném prostoru se šíří pouze přímá vlna, a proto můžeme psát

$$x_j(t) = \sum_{k=1}^N h_{kj} s_k(t), \quad j = 1, \dots, M, \quad (1.1)$$

kde  $x_j(t)$  je pozorovaný signál na  $j$ -tém senzoru,  $s_k(t)$  je  $k$ -tý zdrojový signál,  $h_{kj}$  je váha prostředí od zdroje  $k$  k senzoru  $j$  a  $t$  představuje diskretní čas. V maticovém zápisu se výraz (1.1) zjednoduší na

$$\mathbf{x}(t) = \mathbf{h}\mathbf{s}(t), \quad (1.2)$$

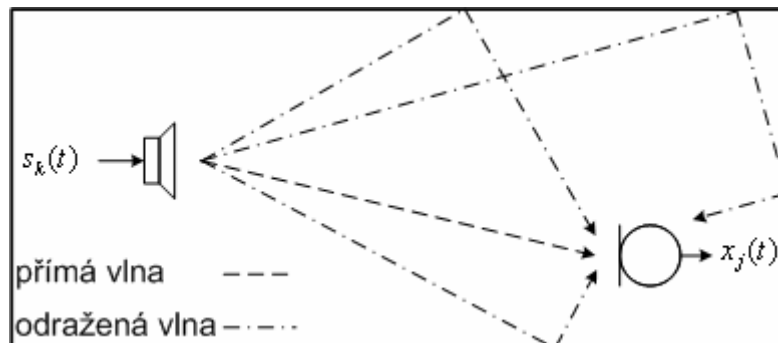
kde  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$  je vektor zdrojových signálů,  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$  je vektor pozorovaných signálů,  $\mathbf{h}$  je směšná matice a  $^T$  reprezentuje transpozici.

Popis směšovacího prostředí pomocí (1.2) označme jako *přímý model*. Přímý model je vhodný pro volné prostředí či bezodrazové komory. V reálných situacích však často bývá zdroj i senzor zvuku umístěn ve vnitřním prostředí, kde kromě přímé vlny vznikají také vlny odražené. Na obr. 1.2 je uvedeno takovéto prostředí s načrtnutými odrazy prvního a druhého řádu, které dopadají na přijímací senzor pod různým úhlem dopadu a s různým zpožděním.

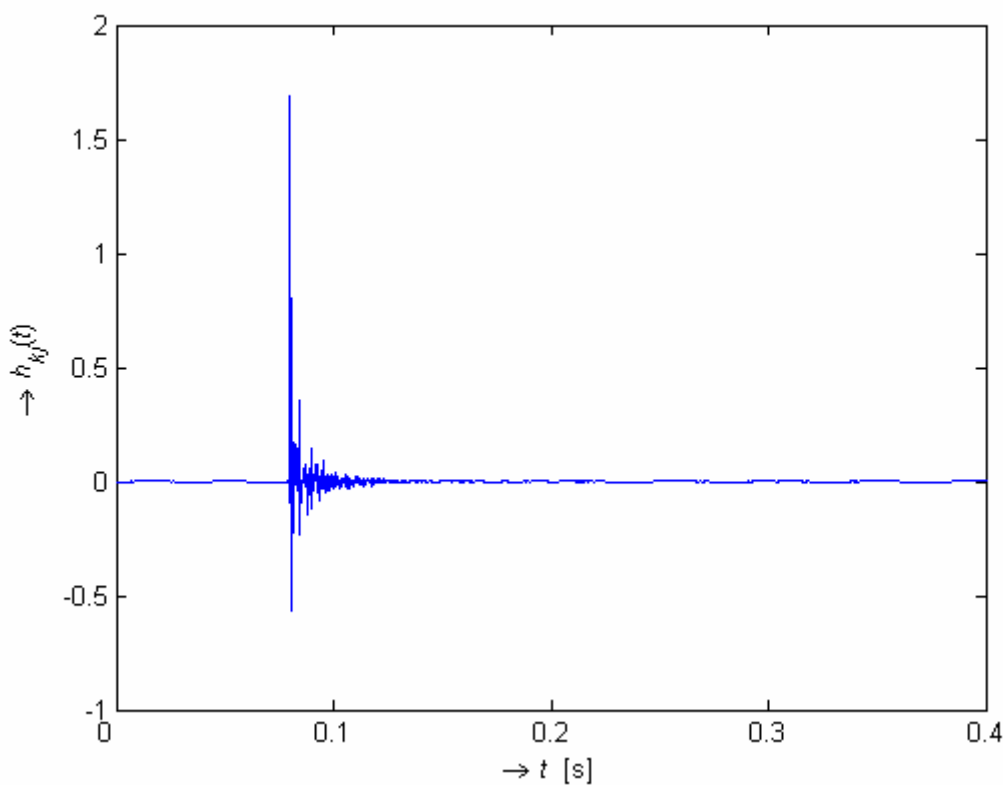
Směšovací prostředí zobrazené na obr. 1.2 budeme označovat jako reálné prostředí a může být popsáno *konvolučním modelem*

$$x_j(t) = \sum_{k=1}^N \sum_{l=0}^{L-1} h_{kj}(l) s_k(t-l), \quad j = 1, \dots, M, \quad (1.3)$$

kde  $h_{kj}(l)$  je impulsní odezva od zdroje  $k$  k senzoru  $j$  a  $L$  je délka impulsní odezvy. Délka impulsní odezvy je vyjádřena jako  $L=T_R f_s$ , kde  $T_R$  je doba dozvuku a  $f_s$  je vzorkovací kmitočet. Ukázka impulsní odezvy je uvedena na obr. 1.3.



Obr. 1.2. Model šíření zvuku v místnosti.



Obr. 1.3. Impulsní odezva s dominantní přímou vlnou – doba dozvuku  $T_R = 120$  ms .

Pokud využijeme krátkodobé Fourierovy transformace (STFT) [15]

$$X_j(f, \tau) = \sum_{t=0}^{T-1} x_j(t) w(t - \tau) e^{-j2\pi f t}, \quad (1.4)$$



kde  $T$  označuje  $T$ -bodovou STFT,  $f \in \left\{0, \frac{1}{T}f_s, \dots, \frac{T-1}{T}f_s\right\}$ ,  $\tau$  představuje označení rámce a  $w(t)$  je váhová posloupnost k zmírnění Gibbsova jevu. Vektor pozorovaných signálů v časově-frekvenční oblasti můžeme zapsat jako

$$\mathbf{X}(f, \tau) = \sum_{k=1}^N \mathbf{H}_k(f) S_k(f, \tau), \quad (1.5)$$

kde  $\mathbf{X}(f, \tau) = [X_1(f, \tau), \dots, X_M(f, \tau)]^T$ , směsný vektor od zdrojového signálu  $k$  ke všem senzorům  $\mathbf{H}_k(f) = [H_{k1}(f), \dots, H_{kM}(f)]^T$  a  $S_k(f, \tau)$  je  $k$ -tý zdrojový signál. V maticovém zápisu se vztah (1.5) zjednoduší na

$$\mathbf{X}(f, \tau) = \mathbf{H}(f) \mathbf{S}(f, \tau), \quad (1.6)$$

kde  $\mathbf{H}(f) = [\mathbf{H}_1(f), \dots, \mathbf{H}_N(f)]$  je směsná matice a  $\mathbf{S}(f, \tau) = [S_1(f, \tau), \dots, S_N(f, \tau)]^T$  je vektor zdrojových signálů.

## 1.2 Separační proces

Úkolem separačního procesu je nalézt separované signály  $y_i(t)$ ,  $i = 1, \dots, N$ , a to *pouze* za použití znalosti vektoru pozorovaných signálů  $\mathbf{x}(t)$  a počtu zdrojových signálů  $N$ . V reálném prostředí je za ideálně separovaný signál možné považovat signál  $y_i(t) = s_k(t)$  resp.  $y_i(t) = \sum_{l=0}^{L-1} h_{kP}(l) s_k(t-l)$ , kde  $P$  je libovolný senzor. V prvním případě je tedy úloha separace rozšířena ještě o dekonvoluci. Samotná slepá dekonvoluce je ovšem obtížná úloha i pouze pro jeden zdrojový signál. Proto se často za separované signály považuje druhý případ, nebo-li zdrojový signál pozorovaný na libovolném senzoru  $P$ .

Při separaci signálů dochází často k permutaci indexů (viz kapitola 2.1) tj. případ kdy  $i \neq k$ . Z důvodu zjednodušení však budeme dále v textu označovat separované signály  $y_k(t)$  shodným indexem jako vstupní signály  $s_k(t)$ .

Proces separace může být proveden v časové nebo časově-frekvenční oblasti. V případě přímého modelu (1.2) je možné provést separaci v časové oblasti. V reálném prostředí, kde vznikají odrazy, je separace v časové oblasti velmi obtížná. Většina BSS algoritmů je navržena pro aplikaci na přímý model. Zjednodušení však nastává s přechodem do časově-frekvenční oblasti, neboť dojde k převodu na přímý model (1.6). Z tohoto důvodu pracuje většina BSS algoritmů v časově-frekvenční oblasti. Nově, již během dokončování této disertační práce, byla představena perspektivní BSS metoda využívající algoritmus EFICA a to přímo v časové oblasti [17].

Volba separačního procesu (separačního algoritmu) závisí na mnoha parametrech jako je např. model směšovacího prostředí, počet zdrojových signálů, počet senzorů. Pro správnou orientaci v současných separačních algoritmech uvedených v kapitole 2 si ve zbytku kapitoly načrtneme stručné rozdělení separačních metod.

### 1.2.1 Jednokanálové metody

Jednokanálové metody mají velkou výhodu v jednoduchosti fyzického uspořádání systému. Výsledný systém je rozměrově malý a finančně nenákladný. Jednokanálové metody pracují v časově-frekvenční oblasti a jsou využívány hlavně k separaci šumu od užitečného signálu. Separace dvou nestacionárních signálů je pouze s jedním senzorem velmi obtížná, neboť postrádáme prostorovou informaci nebo-li možnost odhadu pozice zdroje signálu vůči senzoru. Proto se v případě jednokanálových metod spíše než o separaci hovoří v odborných kruzích o zvýrazňování signálu degradovaného šumem, kde šum je často pokládán za stacionární či po částech stacionární signál.

Základní metodou používanou v oblasti zvýrazňování signálů je spektrální odečítání [22], [39]. Tato metoda se snaží odhadnout rušivý signál, který je následně odečten od pozorovaného signálu. Odhad rušení bývá prováděn v době, kdy zdrojový signál není aktivní. Pro kvalitní zvýraznění zdrojového signálu, musí být systém schopen správně odhadnout začátek a konec aktivity zdrojového signálu, což klade velké nároky na algoritmy detekující aktivitu užitečného signálu (VAD algoritmy) [10].

Pokročilejší separační technikou je metoda RASTA [11], která na základě analýzy spektra eliminuje signály, jež nemají charakter řeči. Pokud je však rušivým signálem řeč je tato metoda neúčinná. Další již robustnější technikou je mapování spektrogramu [26], kde dochází k nalezení masky v časově-frekvenční oblasti, která popisuje aktivitu zdrojového signálu. Násobením masky se spektrogramem dojde k obnově užitečného signálu. Další přístupy jsou založeny také na resyntéze řeči, které detekují parametry řeči jako základní kmitočet, formanty atd. a následně syntetizují již rekonstruovanou řeč [41].

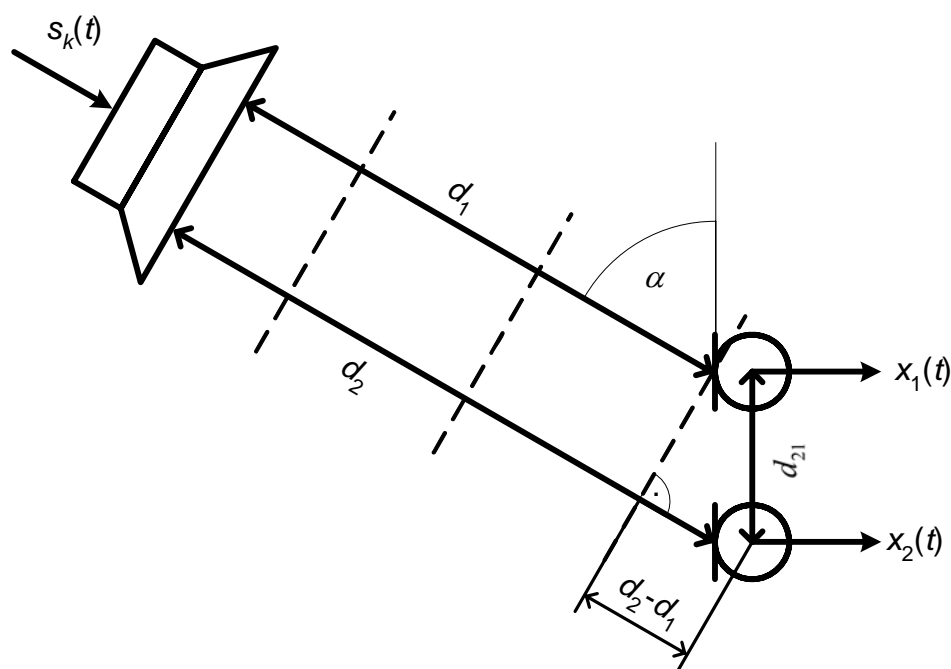
Do této chvíle jsme předpokládali, že používaný senzor je stejně citlivý na signály přicházející ze všech směrů – všesměrový senzor. Prostorovou informaci však můžeme získat použitím směrových senzorů, jejichž citlivost je závislá na úhlu dopadu signálu. Při vhodné pozici senzoru můžeme tedy výrazně potlačit interferující signály a zvýšit účinnost výše uvedených metod. V případě pohybujících se zdrojových signálů či neznámém úhlu dopadu je však použití směrových senzorů nevýhodné neboť nedovolují prostorovou adaptaci.

### 1.2.2 Vícekanálové metody

Vícekanálové metody jsou založeny na zpracování signálů z více senzorů. Senzory tvoří často i celá *senzorová pole* [53]. Pomocí vícekanálových metod jsme samozřejmě schopni dosáhnout kvalitnějších výsledků oproti jednokanálovým metodám, neboť vícekanálové metody poskytují prostorovou informaci a zároveň také umožňují směrovou adaptaci. Hlavní nevýhoda vícekanálových metod spočívala v zvýšené

výpočetní náročnosti. Na současných výpočetních prostředcích, umožňujících paralelní zpracování signálů [30], [54] však není již tento rozdíl dominantní. V současné době je hlavní nevýhodou rozměr sensorového pole a jeho obtížná integrace do mobilních systémů.

Na obr. 1.4 je zdroj rovinného akustického pole s jednoduchým sensorovým polem skládajícím se pouze ze dvou senzorů. Vlnoplochy jsou zobrazeny čárkovanou čarou. Vzdálenost od zdroje k prvnímu senzoru resp. druhému senzoru je označena jako  $d_1$  resp.  $d_2$ . Vzdálenost mezi senzory je  $d_{21}$ . Nositelem prostorové informace je tedy rozdíl vzdáleností  $d_2 - d_1$  určující časový rozdíl mezi dopadem zdrojového signálu na senzory a tím také rozdíl fáze mezi přijatými signály. Úhel  $\alpha$  je označován jako úhel dopadu (DOA) [53].



Obr. 1.4. Příjem rovinného signálu pomocí sensorového pole složeného ze dvou všesměrových senzorů.

#### Geometrie sensorového pole

Důležitou vlastností sensorového pole je vzájemné rozmístění jednotlivých senzorů tzv. *geometrie sensorového pole*. Jednorozměrné sensorové pole (tzv. lineární pole), viz obr. 1.4, nedokáže rozlišit mezi signálem přicházejícím z úhlu  $\alpha$  a  $-\alpha$ , neboť rozdíl vzdáleností  $d_2 - d_1$  a tím i časové zpoždění dopadajících signálů je shodné. V takovémto případě jednorozměrné pole nepředstavuje výhodu oproti jednokanálové metodě. V praktických situacích se tedy jednorozměrná pole používají pro rozsah  $\alpha = \langle 0^\circ, 180^\circ \rangle$ . Dvojměrné sensorové pole je pole umístěné na ploše, které již

umožňuje obsáhnout celý 2D prostor. Dále pak trojrozměrné pole je možné použít k lokalizaci v prostoru.

Geometrické rozmístění senzorů v sensorových polích může být různé. Často používaná pole jsou lineární, trojúhelníková a pravoúhlá popř. i kruhová. Rozmístění senzorů má vliv na přesnost lokalizace zdroje a v různých situacích je výhodnější použít různá uspořádání. Obecné doporučení nejoptimálnějšího rozložení senzorů však neexistuje.

### *Prostorový aliasing*

Návrh sensorového pole vyžaduje brát v úvahu také prostorový aliasing, nebo-li je nutné dodržovat prostorovou vzorkovací poučku

$$d_{\max} < \frac{\lambda_{\min}}{2} = \frac{c}{2f_{\max}}, \quad (1.7)$$

kde  $\lambda_{\min}$  je minimální délka vlny,  $c$  je rychlost zvuku,  $f_{\max}$  je nejvyšší kmitočet složky spektra pozorovaného signálu a  $d_{\max}$  je maximální vzdálenost mezi senzory sensorového pole. Při nedodržení prostorové vzorkovací poučky dochází ke ztrátě jednoznačné závislosti mezi DOA a rozdílem fáze dopadajícího signálu na jednotlivé senzory.

### **1.2.3 Rozdělení podle počtu zdrojů a senzorů**

Volba separační metody závisí kromě počtu senzorů také na počtu zdrojových signálů. Rozlišujeme následující případy:

#### *Určená separace*

Určená separace (DC) nastává v případě, kdy počet zdrojových signálů  $N$  je roven počtu senzorů  $M$ . Základní separační algoritmy jsou založené na metodách lineární prostorové filtrace jako je ICA (kapitola 2.1) [14], [31], BF (kapitola 2.2) [16], [34].

#### *Přeurčená separace*

V případě kdy  $N < M$  hovoříme o přeurčené separaci. Přeurčená separace je řešena převodem na DC a separací metodou ICA (kapitola 2.1) či přímo metodami BF (kapitola 2.2). Převod na DC je prováděn metodou analýza hlavních komponent (PCA) [31]. Cílem PCA je odvodit malý počet hlavních komponent (dekorelovaných lineárních

kombinací), které nesou co nejvíce informací původních proměnných. Jinak řečeno PCA provádí redukci počtu dimenzí.

### *Nedourčená separace*

Nedourčená separace (UDC) je nejsložitější případ vícekanálové separace signálů, kdy počet zdrojových signálů je vyšší než počet senzorů  $N > M$ . V takovémto případě často používaná BSS metoda ICA nemůže být použita. Základními metodami pro řešení UDC jsou nelineární metody předpokládající vzájemnou řidkost zdrojových signálů (kapitola 2.3) [4], [43].

## 2 Přehled současného stavu problematiky

Současné separační metody jsou vícekanálové techniky založené na prostorové filtraci. Separační metody proto můžeme rozdělit podle druhu návrhu prostorových filtrů na analýzu nezávislých komponent ICA, tvarování přijímací charakteristiky BF a metody založené na řídkosti zdrojových signálů SCA.

### 2.1 Analýza nezávislých komponent

Analýza nezávislých komponent (ICA) je statistický přístup k obnově nezávislých signálů z jejich směsí. V teorii ICA jsou nezávislé signály označovány jako nezávislé komponenty. Vraťme se nejdříve ke vztahu (1.6) popisujícímu konvoluční model v časově-frekvenční oblasti. Dále předpokládejme, že směšná matice  $\mathbf{H}(f)$  je známa a  $N = M$ . V takovémto případě je řešení separace signálu triviální

$$\mathbf{Y}(f, \tau) = \mathbf{H}^{-1}(f)\mathbf{X}(f, \tau) = \mathbf{W}(f)\mathbf{X}(f, \tau) = \mathbf{S}(f, \tau), \quad (2.1)$$

kde  $\mathbf{Y}(f, \tau) = [Y_1(f, \tau), \dots, Y_N(f, \tau)]^T$  je vektor separovaných signálů a  $\mathbf{W}(f) = \mathbf{H}^{-1}(f)$  je separační matice. Při slepé separaci signálů však směšovací matice  $\mathbf{H}(f)$  ani separační matice  $\mathbf{W}(f)$  není apriori známa. Jedním ze způsobů, jak nalézt separační matici  $\mathbf{W}(f)$  je právě metoda ICA. Metoda ICA nahlíží na problematiku směšování signálů i následné separace ze statistického pohledu.

#### *Směšovací proces*

Zdrojové signály  $s_k(t)$  jsou nezávislé komponenty, které mohou být popsány vhodným pravděpodobnostním rozdělením. U řečových signálů v časové oblasti pozorujeme relativně dlouhé úseky, kdy signál není aktivní. Můžeme konstatovat, že hustota rozdělení pravděpodobnosti řečového signálu má pozitivní koeficient špičatosti a může být aproximována pomocí Laplaceova rozložení [14], [31].

Naproti tomu pozorované signály  $x_j(t)$  již nejsou vzájemně nezávislé a je možné je aproximovat normálním rozložením. Tato aproximace vychází z centrální limitní věty [31] popisující skutečnost, že součet nezávislých zdrojových signálů se shodným rozložením se blíží normálnímu rozložení [5] mnohem více než samotné zdrojové signály.

Zjednodušeně řečeno, směšovací proces tedy z vzájemně nezávislých signálů  $s_k(t)$  vytváří signály závislé  $x_j(t)$  jejichž hustota rozdělení pravděpodobnosti se blíží normálnímu rozložení.

### *Separační proces*

Separační proces je inverzní proces k směřovacímu procesu. ICA se zde snaží, za využití znalostí ze směřovacího procesu, obnovit z pozorovaných signálů nezávislé komponenty. Jinak řečeno separační proces hledá takovou lineární transformaci (separační matici  $\mathbf{W}$ ), která vede k nalezení maximálně nezávislých komponent, z nichž je pozorovaný signál složen.

K nalezení separační matice je nutné zavést účelovou funkci, která dosahuje extrému právě v případě, kdy dojde k nalezení maximálně nezávislých komponent. Účelová funkce může být odvozena třemi možnými přístupy, a sice na základě vzájemné informace [6], míry popisující odchylku od normálního rozložení [14] což je například koeficient špičatosti [5] a pravděpodobnosti (likelihoodu) [14].

Vzájemnou informaci  $I(\mathbf{y})$  můžeme vyjádřit jako

$$I(\mathbf{y}) = \sum_{k=1}^N H(y_k) - H(\mathbf{y}), \quad (2.2)$$

kde  $H(y_k)$  je marginální entropie,  $H(\mathbf{y})$  je sdružená entropie a  $\mathbf{y} = [y_1, \dots, y_N]^T$  je vektor separovaných proměnných.

Účelovou funkci navrženou na základě vzájemné informace (2.2) můžeme vyjádřit jako [4]

$$\mathfrak{J} = \sum_{k=1}^N H(y_k) - H(\mathbf{x}) - \log|\det \mathbf{W}|. \quad (2.3)$$

Extrém účelové funkce může být nalezen algoritmem největšího spádu

$$\mathbf{W}(q+1) = \mathbf{W}(q) + \eta(q)\Delta\mathbf{W}(q), \quad (2.4)$$

kde  $q$  je iterační krok,  $\eta(q)$  je délka kroku záporného gradientu (rychlost učení) a  $\Delta\mathbf{W}(q)$  je gradient. Vztah pro výpočet gradientu  $\Delta\mathbf{W}(q)$  je

$$\Delta\mathbf{W}(q) = -\frac{\partial \mathfrak{J}}{\partial \mathbf{W}(q)} = (\mathbf{I} - \phi(\mathbf{y}(q))\mathbf{y}^T(q))\mathbf{W}(q), \quad (2.5)$$

kde  $\mathbf{I}$  je jednotková matice,  $\phi(\mathbf{y}(q)) = [\phi(y_1(q)), \dots, \phi(y_N(q))]$  je vektor, jehož každý prvek je roven

$$\phi(y_k(q)) = \frac{\partial \log[p(y_k(q))]}{\partial y_k(q)}, \quad (2.6)$$

kde  $p(y_k(q))$  je model hustoty rozdělení pravděpodobnosti zdrojového signálu.

Navržená separační matice může být s výhodou použita na přímý model směřovacího prostředí přímo v časové oblasti. V případě reálného prostředí je nutné pracovat v časově-frekvenční oblasti, kde je ICA použita na každou časově-frekvenční složku.

S převodem do časově-frekvenční oblasti však nejsou spojena pouze zjednodušení, ale také překážky ve formě *permutace* a *vážení*. Permutace je záměna indexů separovaných signálů oproti zdrojovým signálům. Po provedení ICA máme tedy  $N$  separovaných zdrojů, ovšem časově-frekvenční složky jsou navzájem promíchány. Vážení je spojeno s faktem, že separované signály nemají stejnou energii jako jejich zdrojové protějšky. Překážky ve formě permutace a vážení vznikají při použití ICA na přímý model také v časové oblasti. V tomto případě se však nejedná o podstatný problém. Permutaci a vážení můžeme popsat následujícím vztahem

$$\mathbf{W}(f) = \mathbf{D}(f)\mathbf{P}(f)\mathbf{H}^{-1}(f), \quad (2.7)$$

kde  $\mathbf{P}(f)$  je permutační matice, jež má za následek permutaci indexů separovaných zdrojů a  $\mathbf{D}(f)$  je diagonální matice ovlivňující energii separovaných signálů.

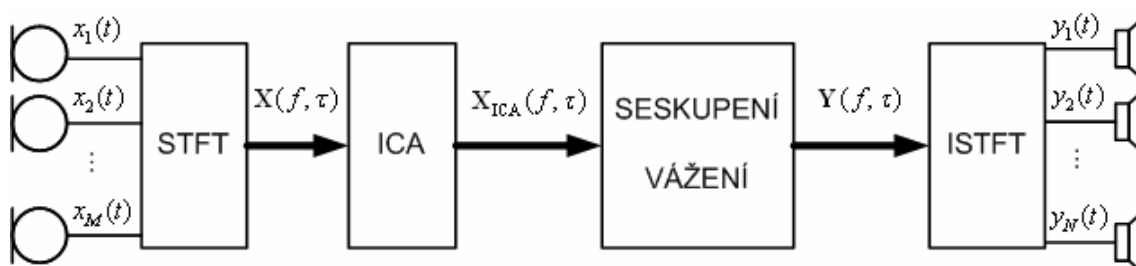
Separační systém na bázi ICA je znázorněn na obr. 2.1. Pozorované signály jsou převedeny pomocí STFT do časově-frekvenční oblasti. Dalším krokem je proces separace metodou ICA. Může být použit jakýkoli algoritmus pracující s komplexními čísly pro přímý model jako např. FastICA [13], [14] JADE [8].

Na tomto místě jsou již signály separovány, ovšem časově-frekvenční složky nejsou správně seskupeny a nemají stejnou energii jako zdrojové signály. Seskupení časově-frekvenčních složek je možné provést např. metodami shlukové analýzy (CA) [20] a vážení např. přizpůsobením na pozorované signály [19]. Výsledný signál je poté převeden pomocí inverzní STFT (ISTFT) [15]

$$y_k(t) = \frac{1}{2\pi w(t-\tau)} \sum_f Y_k(f, \tau) e^{j2\pi f t} \quad (2.8)$$

do časové oblasti.

Pro úplnost ještě uveďme, že v rámci předzpracování signálu bývá zařazeno centrování tj. odstranění stejnosměrné složky a metoda PCA.



Obr. 2.1. Blokový diagram slepé separace pomocí ICA.

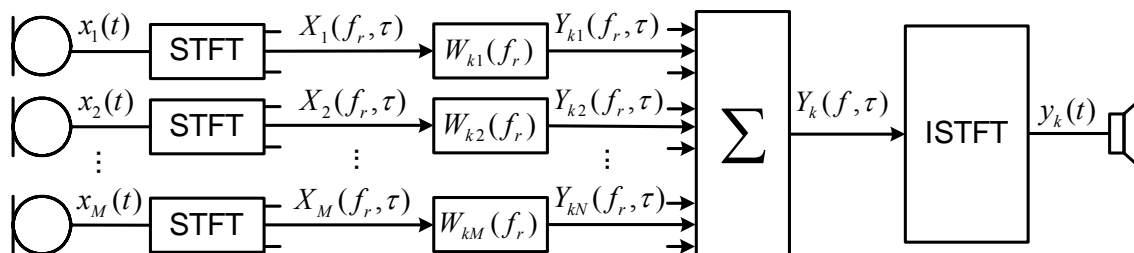


## 2.2 Tvarování přijímací charakteristiky

Tvarování přijímací charakteristiky (BF) je základní technikou prostorové filtrace využívané u senzorových polí. BF umožňuje nastavovat směrovost senzorového pole a tím potlačit interferující signály dopadající z jiného směru než užitečný signál. Blokové schéma tvarovače přijímací charakteristiky (zkráceně tvarovače) pro širokopásmové signály [34], mezi které patří právě řeč či hudba, je uvedeno na obr. 2.2. V každé větvi tvarovače je zařazen filtr s konečnou impulsní charakteristikou (FIR) [37], [51]. Realizace FIR filtru je často prováděna ve frekvenční oblasti, kde  $X_j(f, \tau)$  představuje  $r$ -tou frekvenční složku pozorovaného signálu  $X_j(f, \tau)$ . Každá frekvenční složka je tedy zpracovávána vlastním tvarovačem. Po provedení filtrace jsou jednotlivé frekvenční složky sečteny a separovaný signál

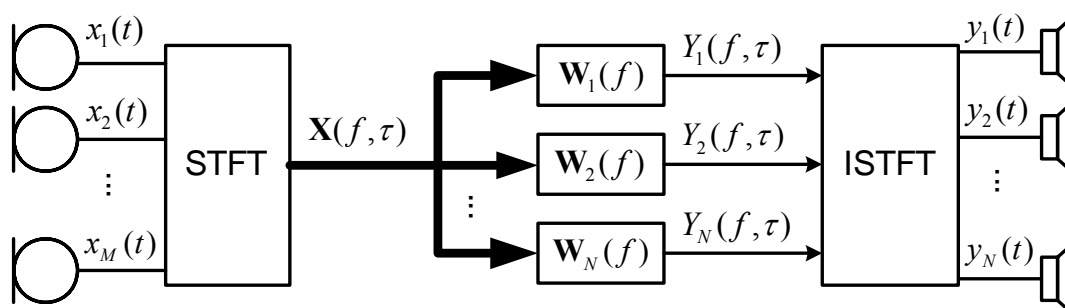
$$Y_k(f, \tau) = \mathbf{W}_k^T(f) \mathbf{X}(f, \tau), \quad (2.9)$$

kde  $\mathbf{W}_k(f) = [W_{k1}(f), \dots, W_{kM}(f)]^T$ , je převeden pomocí ISTFT do časové oblasti.



Obr. 2.2. Blokové schéma tvarovače pro širokopásmové signály v časově-frekvenční oblasti.

Pro realizaci systému schopného separovat  $N$  signálů musí být nasazeno  $N$  tvarovačů, kde  $k$ -tý tvarovač obnovuje  $k$ -tý signál a potlačuje  $N-1$  interferujících signálů. Takovýto systém je znázorněn na obr. 2.3.



Obr. 2.3. Blokové schéma separačního systému složeného z  $N$  tvarovačů.

Metody použitelné k návrhu filtrů separačního systému na obr. 2.3 můžeme rozdělit na datově nezávislé a statisticky optimální. Statisticky optimální metody bývají také označovány jako adaptivní BF.

Datově nezávislé metody, jmenujme např. soufázový tvarovač (Delay and sum beamformer) [16], nulový tvarovač (Null beamformer) [16], navrhuji koeficienty filtru nezávisle na pozorovaných datech. Zjednodušeně řečeno snažíme se, obdobně jako u FIR filtrů, aproximovat požadovanou impulsní odezvu separačního systému. V případě slepé separace signálů není však požadovaná impulsní odezva apriori známa.

Adaptivní metody využívají pozorovaná data k stanovení koeficientů separačních filtrů. Cílem adaptivních metod je navrhnout separační filtry tak, aby separovaný signál obsahoval co nejmenší příspěvek od interferujících signálů dopadajících z jiného směru než požadovaný signál. Statisticky optimální váhy filtrů mohou být navrženy na základě rozdílných kritérií [34]. Z často používaných metod jmenujme vícenásobný potlačovač postraních laloků (MSC), metodu maximálního odstupu signálu od šumu (Max SNR) a metodu lineárně podmíněné minimální variance (LCMV).

Adaptivní metody jsou výhodnější pro návrh separačního systému. Bohužel i v případě adaptivních metod není možné provést návrh tvarovače zcela slepě. Metoda MSC vyžaduje pro návrh separačních filtrů znalost časového úseku, kdy není zdrojový signál aktivní. Metoda Max SNR navíc vyžaduje znalost časového úseku, kdy nejsou interferující signály aktivní. K návrhu metodou LCMV je nutné znát DOA nebo směsný vektor.

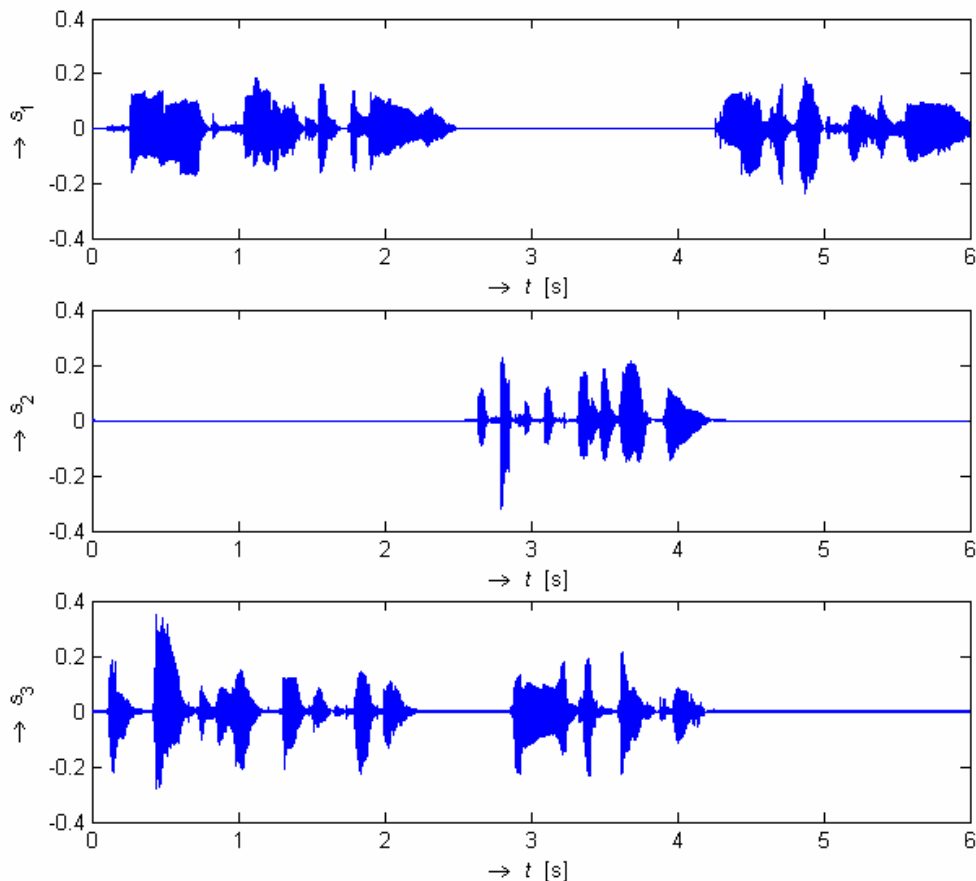
## 2.3 Analýza řídkých komponent v slepé separaci signálů

Analýza řídkých komponent (SCA) je přístup používaný k řešení slepé separace signálů, a to i v případě UDC. SCA klade na zdrojové signály požadavek označovaný jako *řídkost signálu*. Tento požadavek můžeme klasifikovat jako striktnější než je tomu v případě ICA. Řídkost signálu znamená, že signál často nabývá nulové hodnoty či hodnoty blízké nule.

### *Řídkost signálu*

Na obr. 2.4 jsou uvedeny tři promluvy (mužský hlas). Již při letném pohledu na průběhy promluv můžeme konstatovat, že promluvy  $s_1(t)$  a promluva  $s_2(t)$  se vzájemně nepřekrývají (jsou aktivní v různém časovém úseku). V případě smíchání signálů  $s_1(t)$  a  $s_2(t)$  pomocí lineární transformace je možné pouze na základě znalosti aktivity signálu (začátku a konce promluv) provést úspěšnou separaci. V případě

signálů  $s_1(t)$  a  $s_3(t)$  či  $s_2(t)$  a  $s_3(t)$  již takováto separace úspěšná nebude, neboť se signály vzájemně překrývají.



Obr. 2.4. Časový průběh tří promluv – mužský hlas.

Předpoklad pro úspěšnou separaci metodami SCA je minimální překryv zdrojových signálů. Proto SCA uvažuje právě řídké signály, kde je pravděpodobnost vzájemného překrytí daleko menší. V ideálním případě můžeme psát

$$s_k(t) \sum_{p=1, p \neq k}^N s_p(t) = 0, \quad \forall t. \quad (2.10)$$

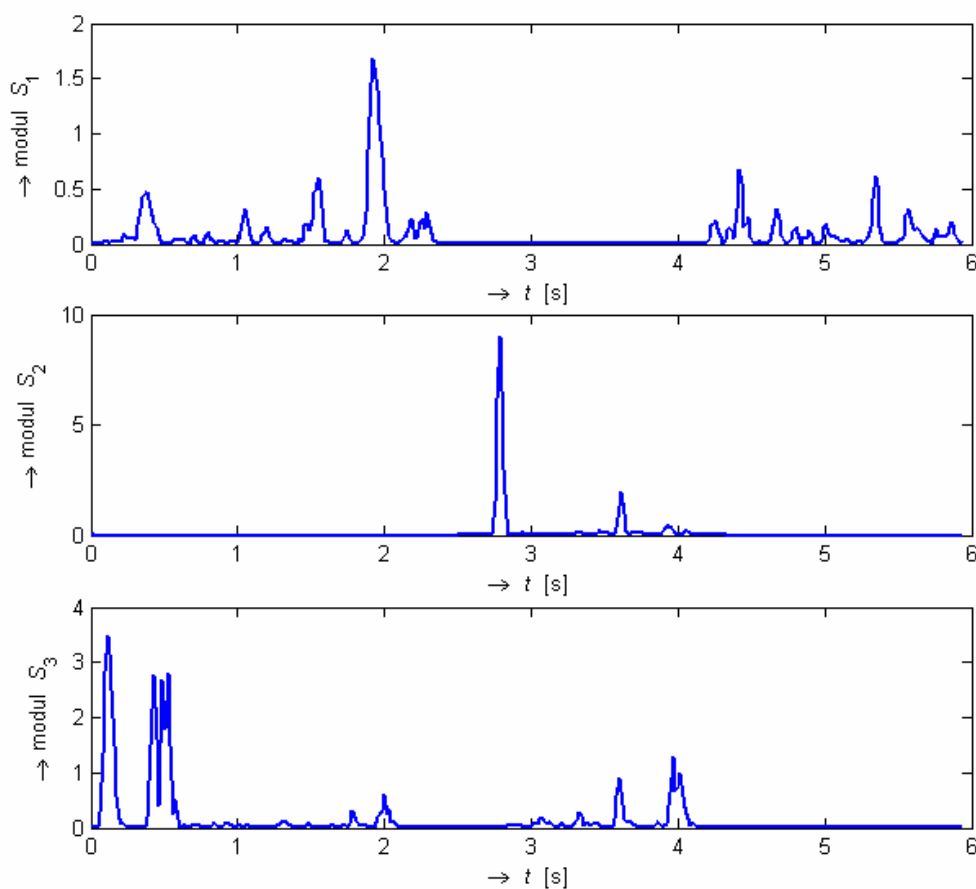
Z obr. 2.4 je zřejmé, že řečové signály v časové oblasti není možné považovat za signály řídké. Vhodným řešením je tedy převod do oblasti, kde vynikne řídkost signálu. Za tímto účelem můžeme využít STFT [7], [42] nebo i vlnkovou transformaci [18], [55].

Pokud převedeme promluvy z obr. 2.4 pomocí STFT do časově-frekvenční oblasti, zjistíme, že časové průběhy modulů promluv  $S_1(f, \tau)$ ,  $S_2(f, \tau)$  a  $S_3(f, \tau)$  se již vzájemně méně překrývají, viz obr. 2.5. Na obr. 2.5 jsou zobrazeny časové průběhy

modulů promluv pro padesátou frekvenční složku ( $f_s = 8$  kHz, 512ti-bodová STFT). Frekvenční složka byla zvolena libovolně. V ideálním případě tedy můžeme psát

$$S_k(f, \tau) \sum_{p=1, p \neq k}^N S_p(f, \tau) = 0, \quad \forall f, \tau. \quad (2.11)$$

V praktických aplikacích nemusí být (2.11) striktně splněno. Pro úspěšné provedení SCA postačí, pokud je vždy jeden ze zdrojových signálů v dané časově-frekvenční složce dominantní tj. má o mnoho větší modul než ostatní signály.



Obr. 2.5. Časové průběhy modulů tří promluv pro padesátou frekvenční složku – mužský hlas.

### Směšování řídkých signálů

Uvažujme směšovací prostředí popsané vztahem (1.5). Impulsní odezvu  $H_{kj}(f)$  můžeme rozepsat na

$$H_{kj}(f) = h_{kj}(1)e^{-j2\pi f T_{kj}(1)} + \sum_{i=2}^I h_{kj}(i)e^{-j2\pi f T_{kj}(i)}, \quad (2.12)$$

kde  $I$  je řád FIR filtru. Dále předpokládáme, že ve směřovacím prostředí dochází pouze k slabým odrazům. V takovém případě může být (2.12) aproximováno pouze přímou vlnou

$$H_{kj}(f) \approx h_{kj} e^{-j2\pi f T_{kj}}. \quad (2.13)$$

Dále uvažujeme směs řídkých signálů s jedním dominantním zdrojem  $S_k(f, \tau)$  v každé časově-frekvenční složce  $(f, \tau)$ . Pozorovaný signál na senzoru  $j$  můžeme zapsat jako

$$X_{kj}(f, \tau) \approx H_{kj}(f) S_k(f, \tau), \quad k \in \{1, \dots, N\}. \quad (2.14)$$

Za výše uvedených předpokladů a ve vektorovém zápisu se vztah popisující reálné prostředí (1.5) zjednoduší na

$$\mathbf{X}(f, \tau) \approx \mathbf{H}_k(f) S_k(f, \tau). \quad (2.15)$$

### Metody SCA

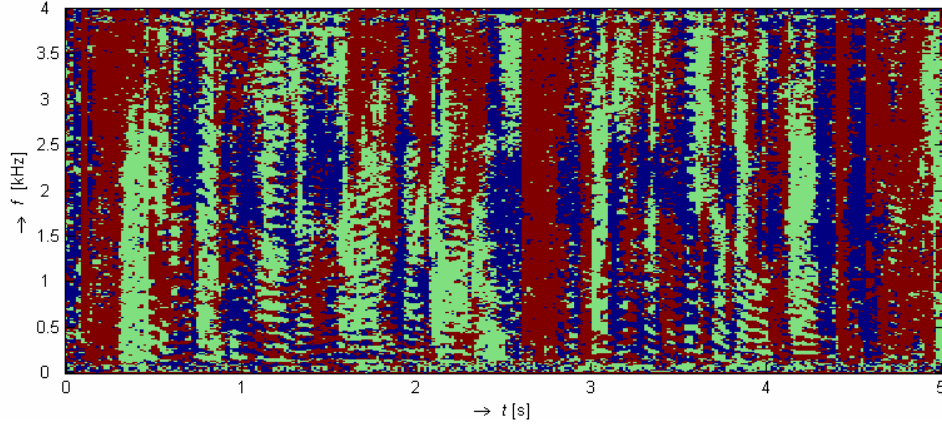
Metody SCA je možné rozdělit do dvou skupin. První skupinou metod jsou tzv. dvoustupňové metody založené na odhadu MAP [18], [35], [36], [42]. Tyto metody v prvním stupni nejprve určí směšnou matici pomocí normalizace a CA (vysvětleno níže v textu). V druhém stupni je provedena vlastní separace.

Druhou skupinou jsou jednostupňové metody, na které se zaměřuje tato disertační práce. Vycházíme ze vztahu (2.15), který uvádí, že v každé časově-frekvenční složce pozorovaného signálu je jeden dominantní zdrojový signál. Tento předpoklad zjednoduší separace signálů na nalezení časově-frekvenčních složek  $(f, \tau)$  v pozorovaném signálu patřících daným zdrojovým signálům, nebo-li na sestavení separační masky.

Vlastní extrakce je provedena pouhým výběrem těchto časově-frekvenčních složek z pozorovaného signálu což bývá realizováno násobením pozorovaného signálu s časově-frekvenční binární maskou [2], [43]. Tento postup budeme nazývat *časově-frekvenční binární maskování* (TFBM). Ukázka separační masky pro pozorovaný signál složený ze směsi tří zdrojových signálů je uvedena na obr. 2.6, kde každá barva odpovídá časově-frekvenčním složkám patřícím jednomu zdrojovému signálu.

Pro určení separační masky předpokládáme prostředí popsané vztahy (2.13) a (2.15) v němž jsou umístěny dva senzory viz obr. 1.4. Poměrem pozorovaných signálů získáme

$$\frac{X_2(f, \tau)}{X_1(f, \tau)} = \frac{H_{k2}(f) S_k(f, \tau)}{H_{k1}(f) S_k(f, \tau)} = \frac{h_{k2} e^{-j2\pi f T_{k2}}}{h_{k1} e^{-j2\pi f T_{k1}}} = \frac{h_{k2}}{h_{k1}} e^{-j2\pi f c^{-1} (d_2 - d_1)}. \quad (2.16)$$



Obr. 2.6. Separační maska – červená barva odpovídá signálu č. 1, zelená barva odpovídá signálu č. 2, modrá barva odpovídá signálu č. 3.

Pokračujeme zavedením *normalizovaného vektoru*

$$\bar{\mathbf{X}}(f, \tau) = \left[ \left| \frac{X_2(f, \tau)}{X_1(f, \tau)} \right|, \arg \left( \frac{X_2(f, \tau)}{X_1(f, \tau)} \right) \right], \quad (2.17)$$

který tvoří základ pro stanovení separační masky. Abychom odstranili závislost na frekvenci [4], upravíme vztah (2.17) na

$$\bar{\mathbf{X}}(f, \tau) = \left[ \left| \frac{X_2(f, \tau)}{X_1(f, \tau)} \right|, \frac{1}{2\pi f} \arg \left( \frac{X_2(f, \tau)}{X_1(f, \tau)} \right) \right]. \quad (2.18)$$

Pro získání DOA dominantního signálu z každé časově-frekvenční složky  $(f, \tau)$  můžeme dále upravit na [3]

$$\bar{\mathbf{X}}(f, \tau) = \left[ \left| \frac{X_2(f, \tau)}{X_1(f, \tau)} \right|, \frac{1}{2\pi f c^{-1} d_{21}} \arg \left( \frac{X_2(f, \tau)}{X_1(f, \tau)} \right) \right], \quad (2.19)$$

kde  $d_{21}$  je vzdálenost mezi senzory, viz obr. 1.4. Dosazením (2.16) do (2.19) získáme

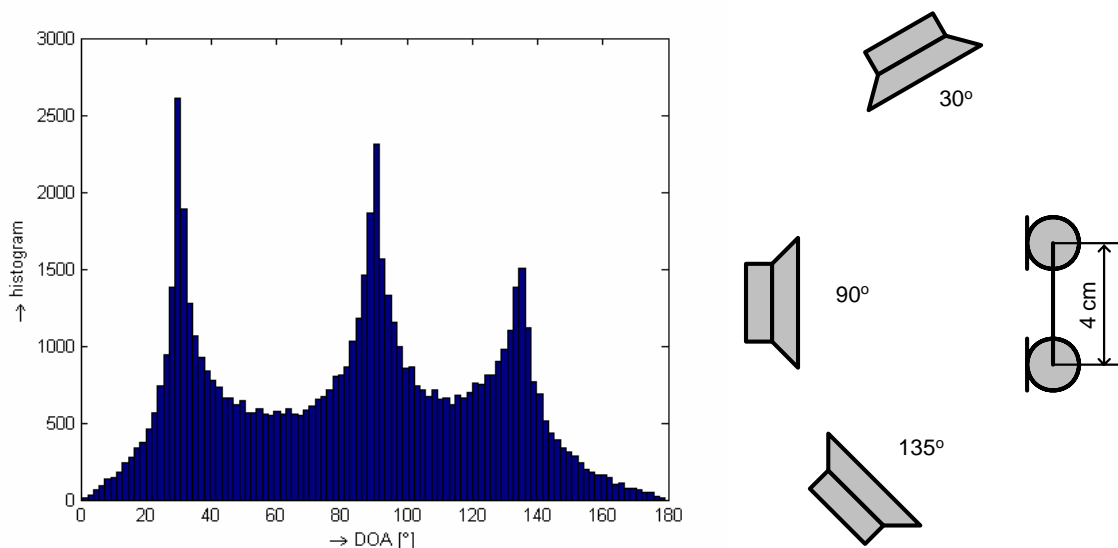
$$\bar{\mathbf{X}}(f, \tau) = \left[ \frac{h_{k2}}{h_{k1}}, \frac{(d_2 - d_1)}{d_{21}} \right]. \quad (2.20)$$

Normalizovaný vektor  $\bar{\mathbf{X}}(f, \tau)$  poskytuje informaci o příslušnosti časově-frekvenčních složek  $(f, \tau)$  k jednotlivým zdrojům v závislosti na pozici zdrojových signálů a sensorového pole. Zjednodušeně řečeno časově-frekvenční složky  $(f, \tau)$  náležející jednomu zdroji tvoří shluky. Pro klasifikaci shluků je možné využít metod CA či pouhého histogramu.

Jako příklad sestavení masky si uveďme vykreslení zjednodušeného normalizovaného vektoru

$$\bar{\mathbf{X}}(f, \tau) = \left[ \frac{1}{2\pi f c^{-1} d_{21}} \arg \left( \frac{X_2(f, \tau)}{X_1(f, \tau)} \right) \right] \quad (2.21)$$

do histogramu, jak je ukázáno na obr. 2.7. Směšování bylo provedeno ve volném prostoru s uvedeným uspořádáním systému. Z histogramu je zřejmé, že pozorovaný signál se skládá ze tří zdrojových signálů. Vrcholy histogramů představují odhad DOA zdrojových signálů. Pro stanovení masky postačí pouze stanovit hranice mezi vrcholy histogramu a tím určit příslušnost časově-frekvenčních složek  $(f, \tau)$  k danému shluku.



Obr. 2.7. Vlevo: histogram normalizovaného vektoru. Vpravo: uspořádání systému.

Obnova zdrojového signálu  $k$  je provedena technikou TFBM nebo-li vynulováním časově-frekvenčních složek nenáležících k shluku  $k$  a ponecháním časově-frekvenčních složek, které náležejí shluku  $k$ .

Nulování ve spektru však kromě odstranění interferujících signálů vnáší do separovaných signálů, obdobně jako spektrální odečítání, krátké tóny, jež jsou náhodně rozloženy ve spektru separovaného signálu. Tento druh rušení je pro posluchače velmi nepřírozený a bývá označován jako *hudební šum*.

## 2.4 Souhrn

V této kapitole byly představeny základní separační systémy. Nejdříve byl popsán separační systém na bázi ICA. Jedná se o často používanou metodu pro separaci signálů v DC uspořádání. Základním požadavkem metody je nezávislost zdrojových signálů. Tento požadavek bývá velmi často splněn. Velkou výhodou je slepý návrh separačního systému a absence nelineárního zkreslení, neboť separace je prováděna lineárními filtry. Nevýhoda se skrývá v omezení pouze na případ kdy  $N \leq M$ . Metoda ICA nemůže být použita v případě, kdy počet zdrojových signálů  $N$  je vyšší než počet

senzorů  $M$  (UDC uspořádání), neboť inverze směsné matice  $\mathbf{H}(f)$  nemá jednoznačné řešení.

Druhý představený separační systém se skládal z  $N$  adaptivních tvarovačů. Každý z tvarovačů představuje lineární prostorový filtr sloužící k obnově jednoho zdrojového signálu  $s_k(t)$ . Separační systém složený z adaptivních tvarovačů může být použitelný dokonce i pro UDC uspořádání. Jak však bude ukázáno v kapitole 5.5 je potlačení interferujících signálů v UDC uspořádání nízké. Základní nedostatek tvarovačů je nemožnost slepého návrhu.

Posledním představeným systémem je jednostupňový separační systém využívající TFBM za předpokladu řídkosti zdrojových signálů. Jedná se ve své podstatě o nelineární prostorovou filtraci založenou na hledání shluků v normalizovaném vektoru  $\bar{\mathbf{X}}(f, \tau)$ . TFBM může být použito dokonce i pro separaci v UDC uspořádání. Návrh systému je možné provést slepě. V separovaných signálech je však často přítomen silný hudební šum, který působí na posluchače velmi negativně.



### 3 Cíle disertační práce

Na základě rozboru provedeném v kapitole 2 bylo zjištěno, že současné separační systémy neumožňují slepý návrh separačního systému, jenž by nedegradoval separované signály hudebním šumem a mohl být použit i v případě, kdy počet zdrojových signálů převyšuje počet senzorů.

V kapitole 2 byly představeny základní přístupy k separaci signálů. Nejdříve byl popsán separační systém na bázi ICA. ICA však nemůže být použita pro nedourčenou separaci. Druhý představený separační systém je založen na použití  $N$  tvarovačů přijímací charakteristiky. Tento systém může být použit dokonce i pro nedourčenou separaci signálů. Návrh systému však není možné provést slepě. Poslední představený separační systém využívá časově-frekvenčního binárního maskování za předpokladu řídkosti zdrojových signálů. Návrh systému je možné provést slepě a to i v případě nedourčeného uspořádání. V separovaných signálech je však často přítomen silný hudební šum.

Cílem této disertační práce je navrhnout systém vícekanálové slepé separace signálů, jež může být použit i v případě nedourčeného uspořádání a zároveň nedegraduje separované signály hudebním šumem. Pro přehlednost shrňme požadavky na separační systém do tří bodů:

- slepý návrh separačního systému,
- použitelnost i pro nedourčenou separaci,
- minimalizace hudebního šumu.

Navržený systém by dále měl splňovat obecné požadavky kladené na systémy slepé separace signálů:

- silné potlačení interferujících signálů,
- malé zkreslení původního užitečného signálu.

Kromě výše uvedených bodů je možné nalézt další kritéria, která by měl splňovat *ideální* separační systém, můžeme jmenovat např. odolnost proti výrazným reverberacím, nízká výpočetní náročnost či realizovatelnost v reálném čase na mobilních zařízeních. Tato práce si však neklade za cíl nalézt ideální separační systém, ale především navrhnout systém, který bude splňovat výše uvedené body a to v prostoru bez výrazných reverberací.

## 4 BSS systémy založené na časově-frekvenčním maskování

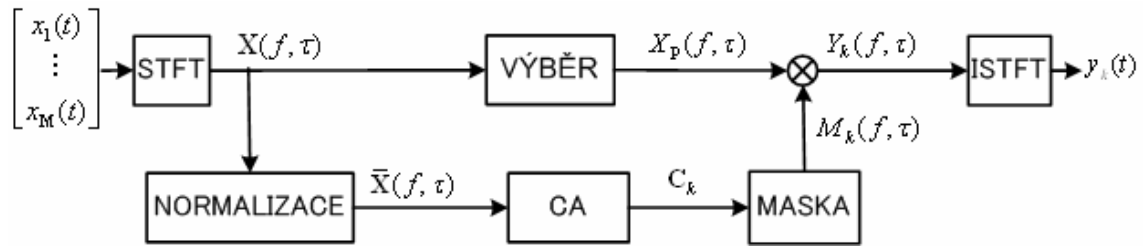
TFBM je přístup umožňující separovat signály i v UDC uspořádání za předpokladu, že zdrojové signály můžeme v časově-frekvenční oblasti označit jako řídké. Jak bylo uvedeno v kapitole 2.3 je základem TFBM separační maska.

Průkopnický přístup v stanovování masky přinesl systémem DUET, který využívá normalizovaných vektorů (2.18), (2.19) [43] a manuální CA. Pozdější přístupy přidali automatické techniky CA [27]. Další přístupy rozšířily DUET z dvoukanálové techniky na vícekanálovou [28].

Zvýšenou robustnost BSS systémů s TFBM přinesl systém MENUET [2]. MENUET může být aplikován na libovolné uspořádání senzorového pole a umožňuje aplikovat standardní metody CA jako např. CA  $k$ -průměrů ( $k$ -means clustering) [9] a hierarchická CA [9].

### 4.1 Systém MENUET

Při návrhu nedourčeného BSS systémů budeme vycházet ze systému MENUET, jehož blokové schéma je uvedeno na obr. 4.1. Vektor pozorovaných signálů  $\mathbf{x}(t)$  je nejdříve převeden pomocí STFT do časově-frekvenční oblasti. Následuje normalizace vektoru  $\mathbf{X}(f, \tau)$ .



Obr. 4.1. Blokové schéma systému MENUET.

MENUET zavádí nový normalizovaný vektor [2]

$$\bar{\mathbf{X}}(f, \tau) = [\bar{\mathbf{X}}^L(f, \tau), \bar{\mathbf{X}}^P(f, \tau)]^T, \quad (4.1)$$

kde

$$\bar{\mathbf{X}}^L(f, \tau) = \left[ \frac{|X_1(f, \tau)|}{\sqrt{\sum_{j=1}^M |X_j(f, \tau)|^2}}, \dots, \frac{|X_M(f, \tau)|}{\sqrt{\sum_{j=1}^M |X_j(f, \tau)|^2}} \right], \quad (4.2)$$

$$\bar{\mathbf{X}}^p(f, \tau) = \left[ \frac{1}{2\pi f c^{-1} d_{\max}} \arg \left( \frac{X_1(f, \tau)}{X_J(f, \tau)} \right), \dots, \frac{1}{2\pi f c^{-1} d_{\max}} \arg \left( \frac{X_M(f, \tau)}{X_J(f, \tau)} \right) \right]. \quad (4.3)$$

$d_{\max}$  je maximální vzdálenost mezi dvěma senzory sensorového pole nebo alespoň její odhad,  $J$  je libovolný senzor, jenž bude označován jako *referenční senzor*.

Přepíšme normalizovaný vektor do výhodnějšího komplexního zápisu ve tvaru

$$\bar{X}_j(f, \tau) = \bar{X}_j^L(f, \tau) \exp \left[ j \bar{X}_j^p(f, \tau) \right]^T, \quad (4.4)$$

kde  $\bar{X}_j^L(f, \tau)$  resp.  $\bar{X}_j^p(f, \tau)$  jsou  $j$ -té komponenty ze vztahu (4.2) resp. (4.3). Normalizaci poté můžeme zapsat jako

$$\begin{aligned} \check{X}_j(f, \tau) &= |X_j(f, \tau)| \exp \left[ j \frac{\arg(X_j(f, \tau) / X_J(f, \tau))}{2\pi f c^{-1} d_{\max}} \right], \\ \bar{\mathbf{X}}(f, \tau) &= \frac{\check{\mathbf{X}}(f, \tau)}{\sqrt{\sum_{j=1}^M |\check{X}_j(f, \tau)|^2}}, \end{aligned} \quad (4.5)$$

kde  $\check{\mathbf{X}}(f, \tau) = [\check{X}_1(f, \tau), \dots, \check{X}_M(f, \tau)]^T$  a  $\bar{\mathbf{X}}(f, \tau) = [\bar{X}_1(f, \tau), \dots, \bar{X}_M(f, \tau)]^T$ .

Normalizovaný vektor  $\bar{\mathbf{X}}(f, \tau)$  je poté rozdělen do  $N$  shluků  $C_1, \dots, C_N$ . Každý z shluků představuje jeden zdrojový signál. Předpokládáme, že počet shluků  $N$  je apriori znám. Vhodným algoritmem z důvodu výpočetní náročnosti je CA  $k$ -průměrů minimalizující účelovou funkci

$$\mathfrak{I} = \sum_{k=1}^N \sum_{\bar{\mathbf{X}}(f, \tau) \in C_k} \delta_k(f, \tau), \quad (4.6)$$

kde  $\delta_k(f, \tau)$  je Euklidovská vzdálenost [9] mezi časově-frekvenčními složkami  $(f, \tau)$  normalizovaného vektoru  $\bar{\mathbf{X}}(f, \tau)$  a *středy shluků*  $\mathbf{c}_k = [c_{k1}, \dots, c_{kM}]^T$ .

Účelová funkce (4.6) je minimalizována iteracemi

$$C_k = \left\{ \bar{\mathbf{X}}(f, \tau) \mid k = \underset{k}{\operatorname{argmin}} (\delta_k(f, \tau)) \right\}, \quad (4.7)$$

$$\mathbf{c}_k = \mathbb{E}[\bar{\mathbf{X}}(f, \tau)]_{\bar{\mathbf{X}}(f, \tau) \in C_k}, \quad (4.8)$$

kde  $\mathbb{E}[\cdot]_{\bar{\mathbf{X}}(f, \tau) \in C_k}$  je střední hodnota časově-frekvenčních složek  $(f, \tau)$  normalizovaného vektoru  $\bar{\mathbf{X}}(f, \tau)$  jež náleží do shluku  $C_k$ .

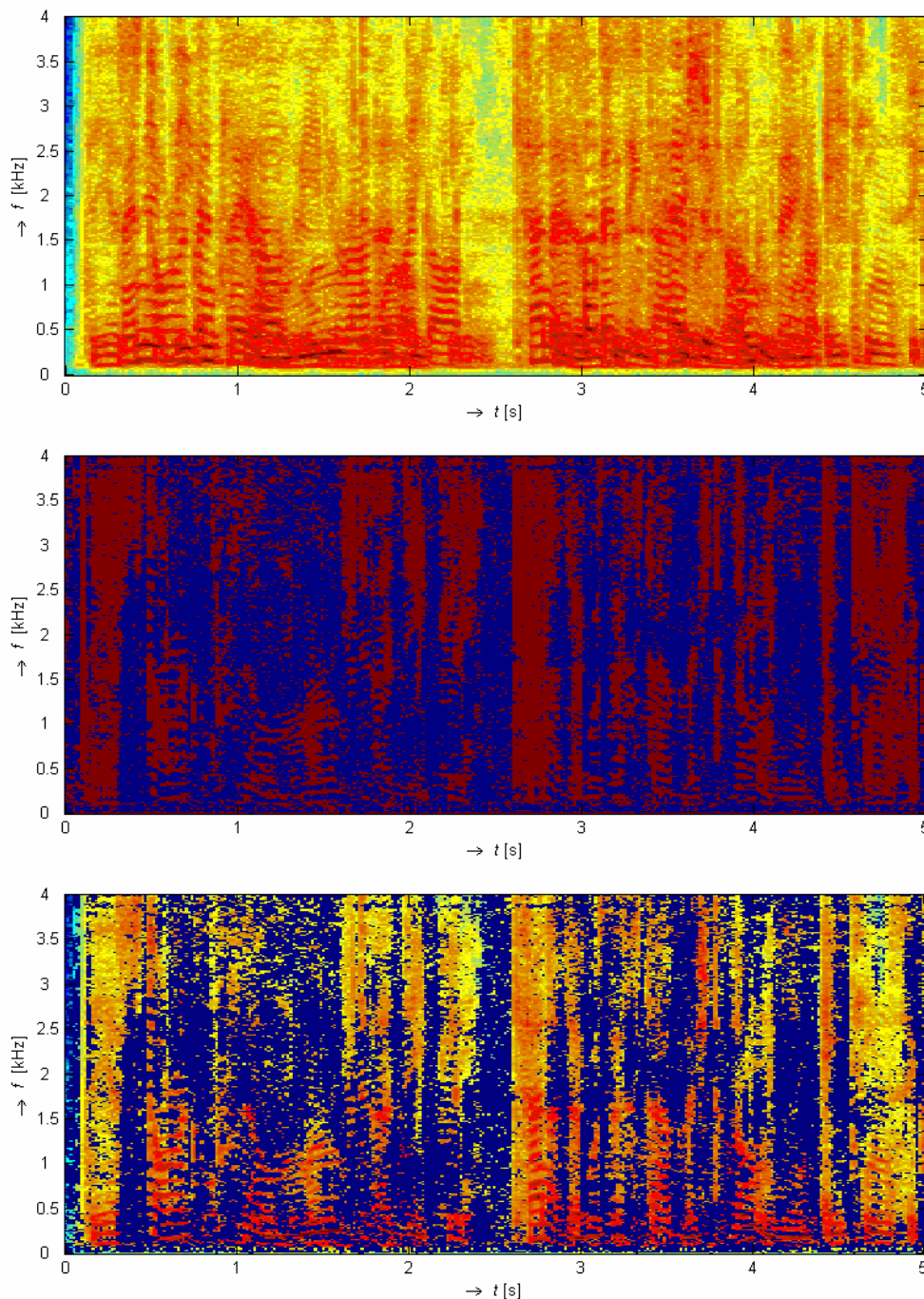
Dalším krokem je sestavení binární masky  $M_k(f, \tau)$  na základě výsledků CA. Binární maska

$$M_k(f, \tau) = \begin{cases} 1 & \bar{\mathbf{X}}(f, \tau) \in C_k \\ 0 & \text{jinak} \end{cases} \quad (4.9)$$

extrahuje všechny prvky vektoru separovaných signálů  $\mathbf{Y}(f, \tau) = [Y_1(f, \tau), \dots, Y_N(f, \tau)]^T$  pouhým násobením s libovolným pozorovaným signálem  $X_p(f, \tau)$ , tedy

$$Y_k(f, \tau) = M_k(f, \tau) X_p(f, \tau), \quad (4.10)$$

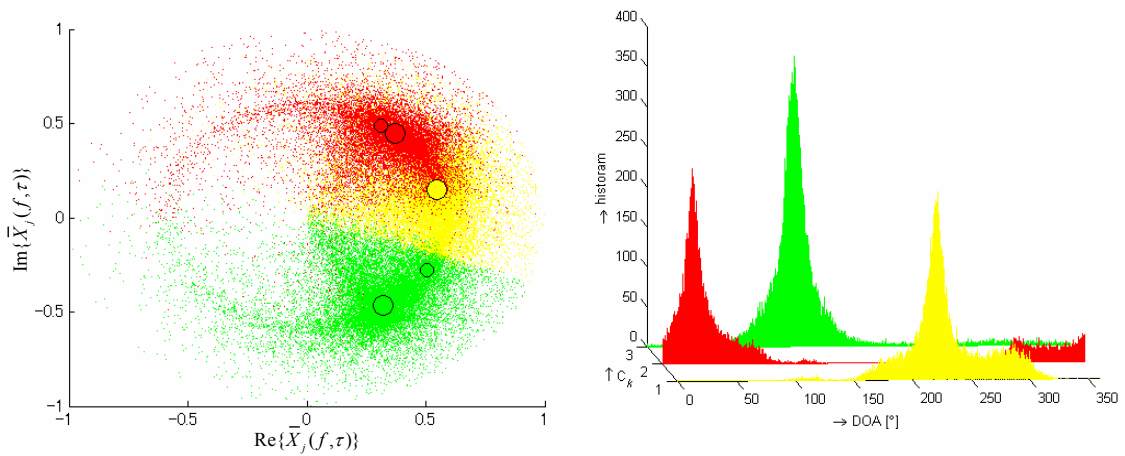
kde  $P \in \{1, \dots, M\}$ . Posledním krokem je převod separovaných signálů  $Y_k(f, \tau)$  do časové oblasti pomocí ISTFT.



Obr. 4.2. Maskování pozorovaného signálu. Nahoře – pozorovaná směs  $N=3$  řečových signálů na senzoru  $j$  v prostoru se slabými reverberacemi. Uprostřed – odhadnutá časově-frekvenční binární maska pro extrakci zdrojového signálu  $k$ . Červená barva je v případě kdy  $M_k(f, \tau) = 1$  a modrá barva když  $M_k(f, \tau) = 0$ . Dole – separovaný signál  $k$ .

Proces maskování v prostoru se slabými reverberacemi je zobrazen na obr. 4.2. Uspořádání experimentálního prostředí je uvedeno na obr. 4.9, kde zdroje zvuku jsou zvýrazněny červenou barvou. Zdrojové signály jsou pěti sekundové promluvy. Byla zvolena vzorkovací frekvence  $f_s = 8 \text{ kHz}$ , 512ti-bodová STFT a posun rámce  $T/4$ .

Pro totožné uspořádání systému si uveďme normalizovaný vektor  $\bar{X}_j(f, \tau)$  viz. obr. 4.3. Barva časově-frekvenční složky  $(f, \tau)$  představuje příslušnost k shluku  $C_k$ . Menší kruh označuje počáteční pozici středu shluku  $c_{jk}$  a větší kruh koncovou pozici dosaženou minimalizací účelové funkce  $\mathfrak{J}$ . Barevná výplň kruhů označuje příslušnost středu shluku  $c_{jk}$  k shluku  $C_k$ . V pravé části obr. 4.3 je představen histogram masky  $M_k(f, \tau)$  v závislosti na DOA.



Obr. 4.3. Vlevo: Normalizovaný vektor  $\bar{X}_j(f, \tau)$ ,  $j \neq J$ , kde barva představuje příslušnost časově-frekvenční složky  $(f, \tau)$  k shluku  $C_k$ . Menší kruh představuje počáteční pozici středu shluku  $c_{kj}$  a větší kruh koncovou pozici středu shluku  $c_{kj}$ , tj. po minimalizaci účelové funkce  $\mathfrak{J}$ . Vpravo: histogram masky  $M_k(f, \tau)$  v závislosti na DOA.

## 4.2 Systém MENUET s více stupni příslušnosti

Separační systém MENUET představený v kapitole 4.1 využívá k extrakci užitečného signálu binární masku, jak je zobrazeno na obr. 4.2. Binární maskování používá pouze dva stupně příslušnosti (4.9) a sice absolutní náležitost či absolutní nenáležitost ke shluku  $C_k$ . MENUET s více stupni příslušnosti [44] je zobecněním systému MENUET. Zavádí více stupňů příslušnosti [49], [52] ke shluku  $C_k$  což vede k návrhu separační masky  $M_k(f, \tau) = \langle 0, 1 \rangle$ .

V této kapitole jsou představeny dva přístupy k návrhu masky s více stupni příslušnosti. MENUET s měkkou maskou (SMMENUET) a MENUET s fuzzy maskou (FMMENUET).

#### 4.2.1 Systém MENUET s měkkou maskou

Separční systém MENUET používá pro stanovení shluků  $C_k$  metodu CA  $k$ -průměrů. Tato metoda však žádným způsobem nepenalizuje složky ležící ve větší vzdálenosti  $\delta_k(f, \tau)$  od středu shluku  $\mathbf{c}_k$ . Odlehlé složky, jež jsou v anglické literatuře označovány jako outliers, jsou promítány do separovaného signálu se stejnou vahou jako složky velmi blízké středu shluku. Vzdálenost od středu shluku však také popisuje odchýlení časově-frekvenční složky od odhadnuté DOA.

Systém SMMENUET zavádí tedy penalizaci pro odlehlé složky ve formě nižšího stupně příslušnosti k shluku  $C_k$  čímž je možné dosáhnout zvýšení směrovosti. Zvýšení směrovosti je spojeno s vyšším potlačením interferujících časově-frekvenčních složek, které mohou také způsobovat hudební šum.

Zavedme masku [29], jež povoluje více stupňů příslušnosti a zároveň penalizuje odlehlé složky

$$M_k(f, \tau) = \begin{cases} \frac{1}{1 + e^{g(\theta_k - \theta_T)}} & \bar{\mathbf{X}}(f, \tau) \in C_k, \\ 0 & \text{jinak} \end{cases}, \quad (4.11)$$

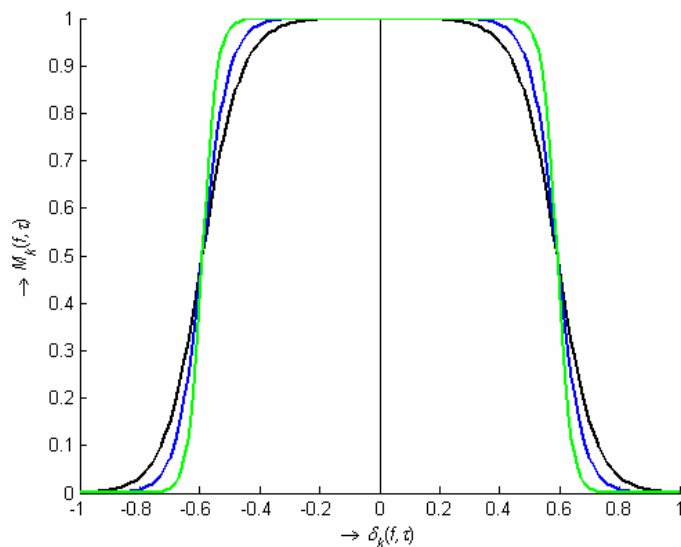
kde

$$\theta_k = \arccos\left(1 - \frac{\delta_k(f, \tau)^2}{2}\right), \quad (4.12)$$

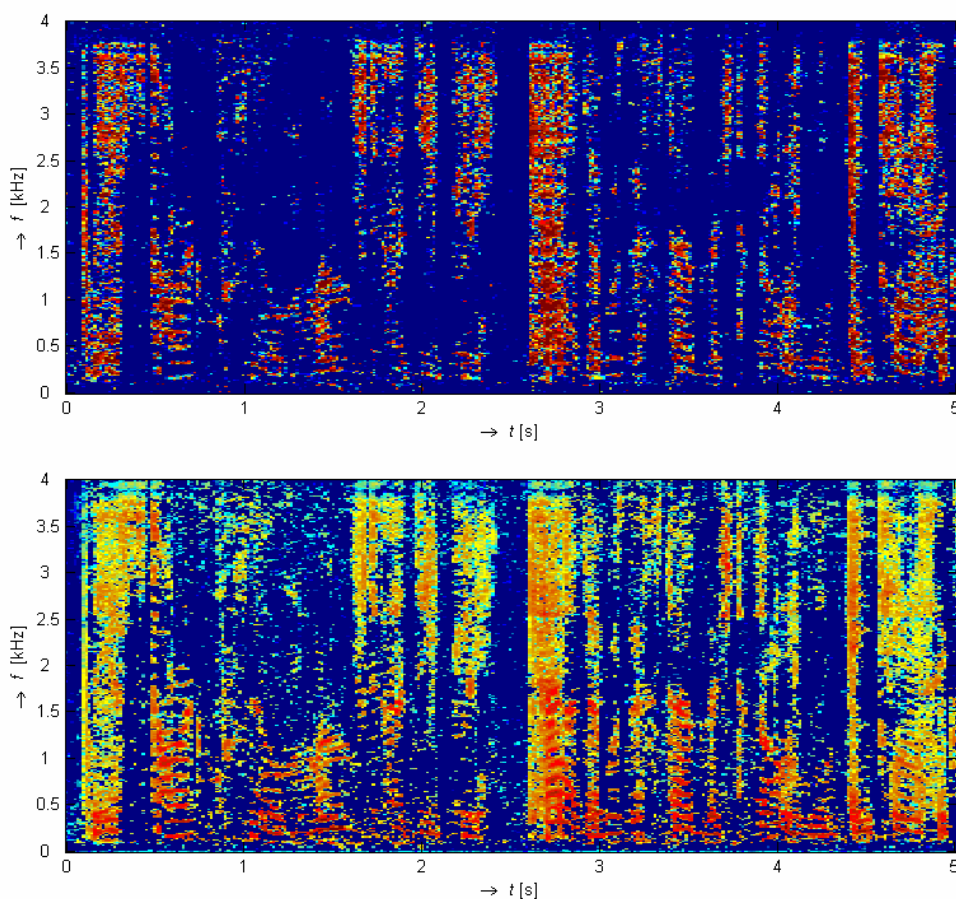
$g$  je strmost přechodu a  $\theta_T$  je bod přechodu. Na obr. 4.4 je zobrazen stupeň příslušnosti v závislosti na vzdálenosti  $\delta_k(f, \tau)$ .

Blokové schéma systému SMMENUET je obdobné se schématem na obr. 4.1. Úprava spočívá v nahrazení vztahu (4.9) vztahem (4.11). Navržená separační maska a separovaný signál je zobrazen na obr. 4.5. Maska byla navržena pro totožné signály v totožném prostředí jako výsledky uvedené na obr. 4.2.

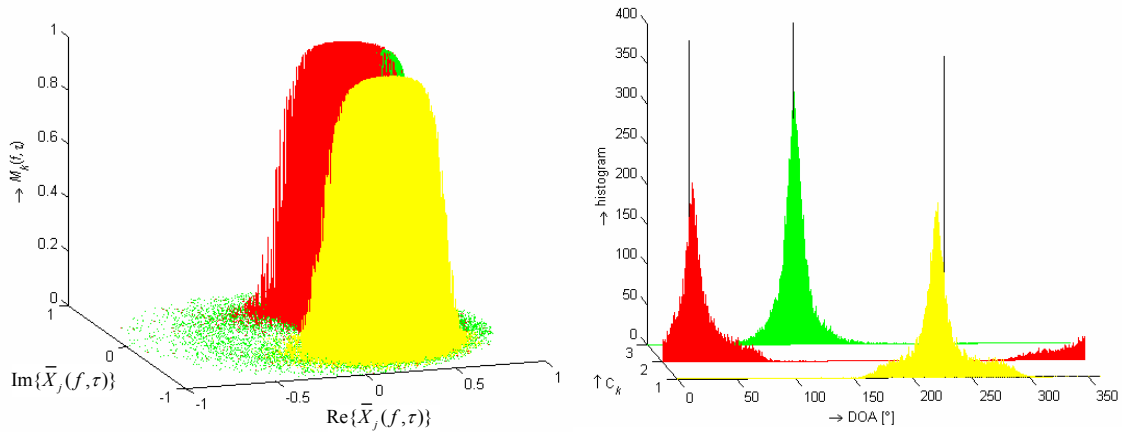
Normalizovaný vektor  $\bar{\mathbf{X}}_j(f, \tau)$  společně s maskou  $M_k(f, \tau)$  a histogram masky  $M_k(f, \tau)$  v závislosti na DOA jsou uvedeny na obr. 4.6. Příslušnost časově-frekvenčních složek  $(f, \tau)$  ke shluku  $C_k$  je vyznačena barevně.



Obr. 4.4. Stupeň příslušnosti k shluku  $C_k$  v závislosti na vzdálenosti  $\delta_k(f, \tau)$  od středu shluku  $c_k$  pro  $\theta_T = 0,6$ . Černá:  $g = 15$ , modrá:  $g = 22$ , zelená:  $g = 40$ .



Obr. 4.5. Maskování pozorovaného signálu. Nahoře – odhadnutá časově-frekvenční maska pro extrakci zdrojového signálu  $k$  systémem SMMENUET. Červená resp. modrá barva představuje vysoký resp. nízký stupeň příslušnosti k shluku  $C_k$ . Dole – separovaný signál  $k$ .



Obr. 4.6. Vlevo: Normalizovaný vektor  $\bar{X}_j(f, \tau)$ ,  $j \neq J$  společně s maskou  $M_k(f, \tau)$  navrženou systémem SMMENUET při  $g = 40$ ,  $\theta_T = 0,6$ . Barva představuje náležitost časově-frekvenční složky  $(f, \tau)$  ke shluku  $C_k$ . Vpravo: histogram masky  $M_k(f, \tau)$  v závislosti na DOA.

#### 4.2.2 Systém MENUET s fuzzy maskou

FMMENUET využívá namísto klasického CA  $k$ -průměrů zobecněnou fuzzy CA  $k$ -průměrů (fuzzy  $k$ -means clustering) [9]. Tato metoda stanovuje příslušnost ke shluku  $C_k$  již během procesu minimalizace účelové funkce. Příslušnost je normována dle

$$\sum_{k=1}^N M_k(f_r, \tau_u) = 1, \quad (4.13)$$

kde  $(f_r, \tau_u)$  je  $r$ -tá frekvenční složka v  $u$ -tém časovém rámci. Fuzzy CA  $k$ -průměrů hledá minimum účelové funkce

$$\mathfrak{J} = \sum_{k=1}^N \sum_{\mathbf{x}(f, \tau)} M_k^b(f, \tau) \delta_k(f, \tau), \quad (4.14)$$

pomocí iterací

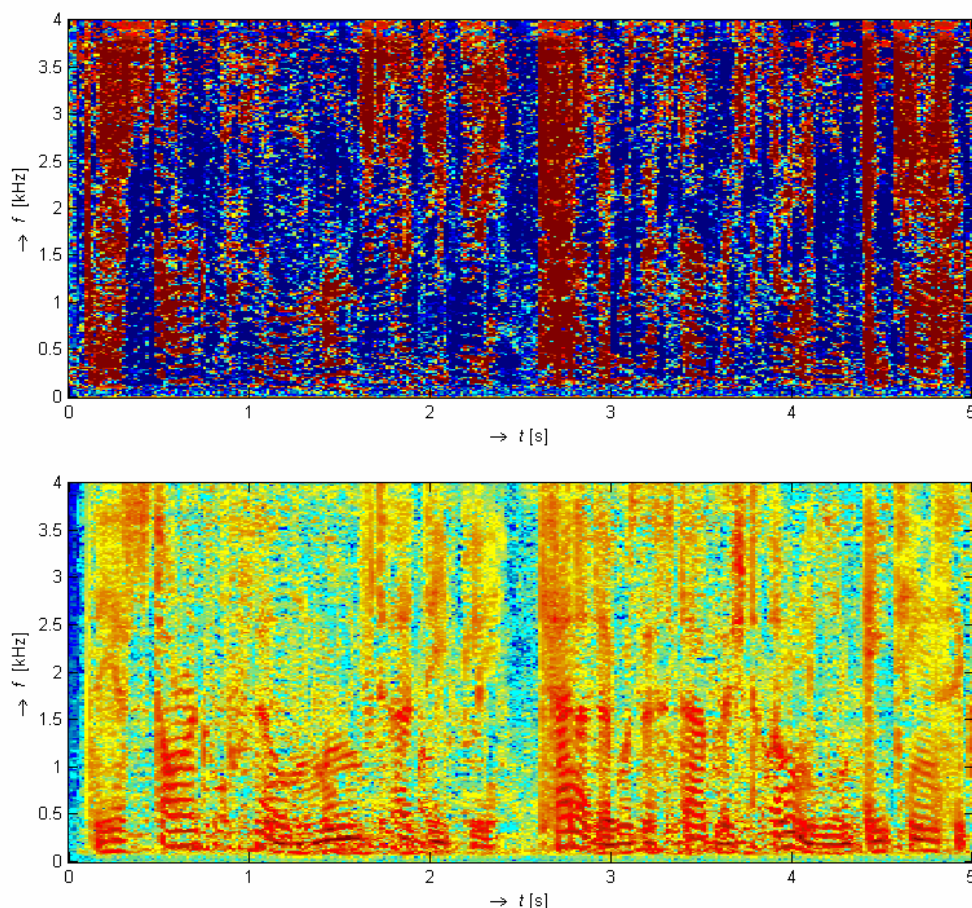
$$c_{kj} = E \left[ \frac{M_k(f, \tau) \bar{X}_j(f, \tau)}{M_k(f, \tau)} \right], \quad M_k(f, \tau) = \frac{\delta_k(f, \tau)^{\frac{1}{b-1}}}{\sum_{k=1}^N \delta_k(f, \tau)^{\frac{1}{b-1}}}, \quad (4.15)$$

kde  $E[\cdot]$  je střední hodnota a  $b > 1$  je parametr určující vzájemné prolínání, tzv. blending, jednotlivých shluků.

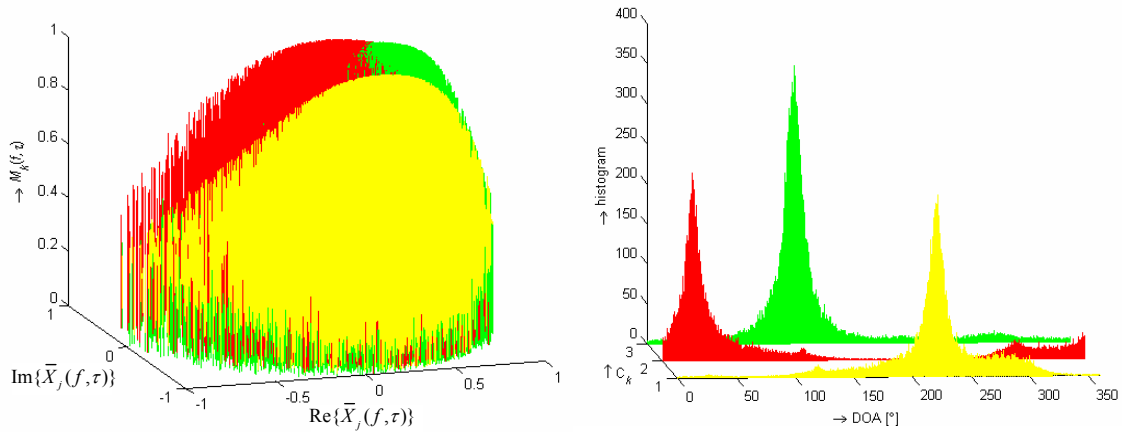
Fuzzy CA  $k$ -shluků dovoluje omezit nulování časově-frekvenčních složek. Nula je časově-frekvenční složce přiřazena velmi zřídka, čímž je možné dosáhnout omezení hudebního šumu. Nežádoucím jevem je poté zvýšené prosakování interferencí do užitečného signálu.



Separační maska navržená systémem FMMENUET je zobrazena na obr. 4.7. Maska byla navržena pro totožné systémové uspořádání jako maska na obr. 4.2. Normalizovaný vektor  $\bar{X}_j(f, \tau)$  společně s maskou  $M_k(f, \tau)$  a histogramem masky  $M_k(f, \tau)$  v závislosti na DOA jsou uvedeny na obr. 4.8. Příslušnost časově-frekvenčních složek  $(f, \tau)$  k shluku  $C_k$  je vyznačena barevně.



Obr. 4.7. Maskování pozorovaného signálu. Nahoře – odhadnutá časově-frekvenční maska pro extrakci zdrojového signálu  $k$  systémem FMMENUET. Červená resp. modrá barva představuje vysoký resp. nízký stupeň příslušnosti k shluku  $C_k$ . Dole – separovaný signál  $k$ .



Obr. 4.8. Vlevo: Normalizovaný vektor  $\bar{X}_j(f, \tau)$ ,  $j \neq J$  společně s maskou  $M_k(f, \tau)$  navrženou systémem FMMENUET. Barva představuje příslušnost časově-frekvenční složky  $(f, \tau)$  ke shluku  $C_k$ . Vpravo: histogram masky  $M_k(f, \tau)$  v závislosti na DOA.

### 4.3 Experimentální část

V této části prověříme představené BSS systémy. Vyhodnocení budeme provádět na separovaných signálech, a to pomocí poměru užitečného signálu vůči interferujícím signálům od ostatních zdrojů

$$\text{SIR}_{\text{OUT}}^k = 10 \log \frac{\text{E}[y_{kk}(t)^2]}{\text{E}\left[\sum_{\bar{b}=1, \bar{b} \neq k}^N y_{k\bar{b}}(t)^2\right]} \quad [\text{dB}], \quad (4.16)$$

kde  $y_{k\bar{b}}(t)$  je  $\bar{b}$ -tý interferující zdroj obsažený v  $k$ -tém separovaném signálu,  $y_{kk}(t)$  je  $k$ -tý výstupní signál bez příspěvku interferujících zdrojů. Kromě  $\text{SIR}_{\text{OUT}}^k$  budeme využívat míru popisující zkreslení užitečného signálu BSS systémem

$$\text{SDR}^k = 10 \log \frac{\text{E}[x_{pk}(t)^2]}{\text{E}\left[(x_{pk}(t) - \gamma y_{kk}(t - \Delta))^2\right]} \quad [\text{dB}], \quad (4.17)$$

kde  $x_{pk}(t) = \sum_{l=0}^{L-1} h_{kp}(l)s_k(t-l)$  a koeficienty  $\gamma$  resp.  $\Delta$  kompenzují změnu amplitudy resp. zpoždění zavedené BSS systémem.

Pro zjištění poměru užitečného signálu vůči interferencím od ostatních zdrojů na vstupu BSS systému využijeme

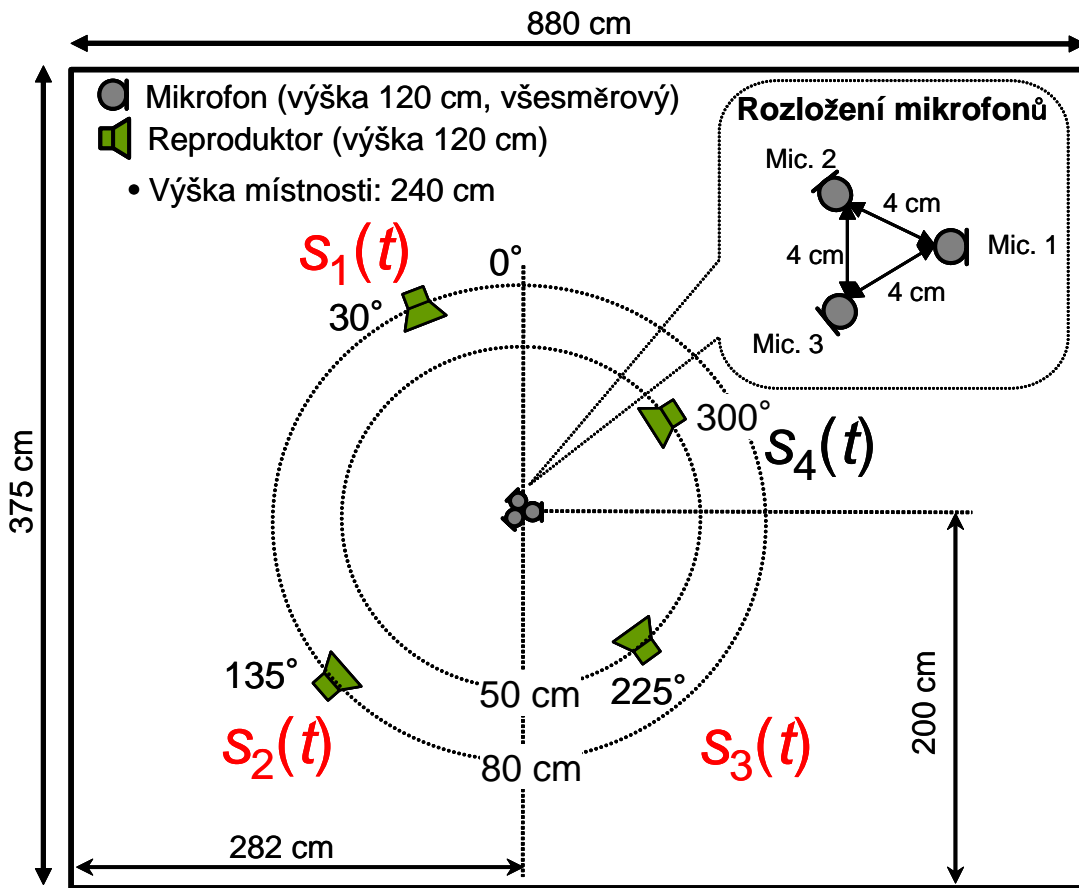
$$\text{SIR}_{\text{IN}}^k = 10 \log \frac{\text{E}[x_{pk}(t)^2]}{\text{E}\left[\sum_{\bar{b}=1, \bar{b} \neq k}^N x_{p\bar{b}}(t)^2\right]} \quad [\text{dB}]. \quad (4.18)$$

Potlačení interferencí BSS systémem pro  $k$ -tý signál určíme z

$$\text{SIR}^k = \text{SIR}_{\text{OUT}}^k - \text{SIR}_{\text{IN}}^k. \quad (4.19)$$

Dále zavedme aritmetický průměr  $\mu$  a směrodatnou odchylku  $\sigma$  souboru hodnot  $\text{SIR}^k$  a  $\text{SDR}^k$ . Soubor hodnot bude u každého experimentu specifikován. Pro zjednodušení budeme využívat dále pouze výrazů SIR a SDR.

Experimenty byly prováděny v prostředí zobrazeném na obr. 4.9. Při DC uspořádání  $M=N=3$  byly použity pouze zdrojové signály, jež jsou zvýrazněny červenou barvou. Během UDC uspořádání  $M=3, N=4$  byl navíc zahrnut zdrojový signál s  $\text{DOA}=300^\circ$ . Jako zdrojové signály byly využity mužské a ženské promluvy v českém jazyce v délce 5 sekund. Byla použita vzorkovací frekvence  $f_s = 8 \text{ kHz}$  a 512ti-bodová STFT. Rámec byl posouván o 1/4 své délky. Experimenty byly provedeny v jazyce Matlab, kde byly také zdrojové signály vzájemně smíchány podle vztahu (1.3). Jedna z impulsních odezev, jež byly během směřovacího procesu využity, je zobrazena na obr. 1.3.



Obr. 4.9. Uspořádání experimentálního systému.

Výsledky experimentu jsou shrnuty v tab. 4.1. Soubor hodnot SIR a SDR se skládá z  $N$  zdrojových signálů všech možných kombinací mužských a ženských promluv. Tedy např. v DC uspořádání jsou hodnoty v tab. 4.1 vypočteny ze čtyř

experimentů {muž, muž, muž}, {muž, muž, žena}, {muž, žena, žena}, {žena, žena, žena}. Strmost přechodu a bod přechodu resp. prolínání u systému SMMENUET resp. FMMENUET jsou v tab. 4.1 označeny jako SMMENUET ( $g; \theta_T$ ) resp. FMMENUET ( $b$ ).

Tab. 4.1. Výsledky experimentů.

	DC				UDC			
	SIR [dB]		SDR [dB]		SIR [dB]		SDR [dB]	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
MENUET	15,1	1,5	11,4	1,4	16,2	0,9	9,9	0,7
SMMENUET (15;0,3)	21,2	2,0	6,7	0,8	21,1	1,0	6,7	0,6
SMMENUET (22;0,3)	22,0	1,9	6,1	0,7	21,8	0,9	6,3	0,5
SMMENUET (40;0,3)	22,5	1,8	5,6	0,7	22,2	0,9	6,0	0,5
SMMENUET (15;0,6)	17,3	1,8	9,4	1,0	17,8	0,9	8,7	0,7
SMMENUET (22;0,6)	17,2	1,8	9,4	1,0	17,6	0,9	8,8	0,7
SMMENUET (40;0,6)	17,0	1,8	9,5	1,0	17,4	0,9	8,8	0,7
SMMENUET (15;0,9)	15,3	1,5	11,1	1,3	16,3	0,9	9,7	0,8
SMMENUET (22;0,9)	15,4	1,5	11,2	1,3	16,4	0,9	9,7	0,8
SMMENUET (40;0,9)	15,4	1,5	11,2	1,3	16,4	0,9	9,7	0,8
FMMENUET (1,2)	15,5	1,5	11,6	1,4	16,6	0,7	10,1	0,7
FMMENUET (1,5)	15,8	1,5	11,7	1,3	17,0	0,7	10,1	0,7
FMMENUET (1,8)	15,6	1,4	11,6	1,2	16,6	0,7	9,9	0,6

Z tab. 4.1 je vidět, že BSS systémy ve většině případů dosahují vyšších hodnot SIR v UDC než v DC upořádání, což může být překvapující. Je to způsobeno strmějším poklesem hodnoty  $SIR_{IN}^k$  než hodnoty  $SIR_{OUT}^k$  při přechodu do UDC uspořádání. Pokud bychom brali v úvahu pouze  $SIR_{OUT}^k$  místo celkového potlačení interferencí podle vztahu (4.19) bylo by dosaženo nižších hodnot. V UDC uspořádání je také dosahováno menší směrodatné odchylky jak pro SIR tak SDR.

Z tab. 4.1 je dále zřejmé, že u systému SMMENUET, dochází při přibližování bodu přechodu  $\theta_T$  k středu shluků k zvyšování SIR a poklesu SDR. Zvyšováním strmosti přechodu  $g$  dochází k zvyšování SIR a poklesu SDR, a to pouze pokud je bod

přechodu relativně blízko k středu shluku. V opačném případě dochází k poklesu SIR a zvyšování SDR. Systém SMMENUET tedy umožňuje přizpůsobovat BSS požadavkům na velikost potlačení interferencí za cenu zvýšení zkreslení, jež je vnášeno do separovaného signálu. K současnému zlepšení obou parametrů SIR a SDR oproti systému MENUET však nedochází.

Systém FMMENUET zlepšuje oproti systému MENUET jak SIR tak i SDR. Je možné nalézt koeficient prolínání  $b$ , při němž je dosaženo maximálních hodnot SIR a SDR. V uvedeném experimentu se jednalo o  $b=1,5$ . Zlepšení systému však není významného charakteru a je vykoupeno zvýšenými výpočetními nároky.

Míru degradace signálu hudebním šumem není možné určit na základě SIR a SDR. Míru hudebního šumu v separovaném signálu je možné stanovit pomocí subjektivního testu, viz kapitola 8.

#### 4.4 Souhrn

BSS systémy představené v této kapitole jsou založeny na systému MENUET. Zavádějí více stupňů příslušnosti než pouhou absolutní příslušnost a absolutní nepřislušnost. SMMENUET umožňuje zvýšit směrovost, jak je ukázáno na obr. 4.6. Tím dochází k vyššímu potlačení interferencí, ale bohužel také k zvýšené míře zkreslení zaváděné BSS systémem do separovaného signálu. FMMENUET dosahuje lepších výsledků než systém MENUET, viz tab. 4.1, což je způsobeno zavedením více stupňů příslušnosti, již během procesu CA. Je možné najít koeficient prolínání  $b$  při němž je dosaženo maximálních hodnot SIR a SDR. Nastavení koeficientu prolínání  $b$  se však může lišit pro různá systémová uspořádání.

I přes zavedení separační masky s více stupni příslušnosti není možné hudební šum zcela odstranit. Dochází stále k nelineárním změnám ve spektru signálu, které mohou v menší či větší míře degradovat separované signály hudebním šumem.

## 5 Kombinace tvarovačů přijímací charakteristiky a systému MENUET

Systém MENUET je založen na nelineárním přístupu k nedourčené BSS předpokládající vzájemnou řídkost zdrojových signálů. Separované signály jsou extrahovány pomocí binární masky vnášející do separovaných signálů hudební šum. Zavedením masky s více stupni příslušnosti, viz SMMENUET a FMMENUET není možné zcela odstranit hudební šum, neboť extrakce užitečného signálu je stále prováděna separační maskou.

Hudební šum může být odstraněn použitím lineárních metod prostorové filtrace jako je např. BF. Tyto metody však neumožňují slepou separaci. Jako vhodná cesta se tedy jeví kombinace lineární prostorové filtrace a systému časově-frekvenčního maskování, kdy systém MENUET může být využit k určení neznámých veličin, nutných pro slepý návrh lineárních prostorových filtrů [46].

Musíme tedy zvolit takové metody návrhu prostorových filtrů, které vyžadují pouze takové neznámé veličiny, jež můžeme určit systémem MENUET. Dále je nutné výsledné tvarovače adaptovat na aktuální situaci v separačním prostředí tak, aby bylo možné potlačit i nestacionární rušivé signály. Jako vhodné se tedy jeví statisticky optimální metody BF. Dále v této kapitole si proto blíže popíšeme Max SNR tvarovač [33], [34] a LCMV tvarovač [33], [34].

### 5.1 Tvarovač přijímací charakteristiky typu Max SNR

Tvarovač přijímací charakteristiky typu Max SNR (zkráceně Max SNR tvarovač) je statisticky optimální metoda BF při níž se snažíme maximalizovat poměr mezi energií užitečného signálu a energií interferujících signálů na výstupu prostorového filtru

$$\varepsilon(f) = \frac{E[Y_k(f, \tau)Y_k^*(f, \tau)]_{P_T^k}}{E[Y_k(f, \tau)Y_k^*(f, \tau)]_{P_N^k}} = \frac{\mathbf{W}_k^H(f)E[\mathbf{X}(f, \tau)\mathbf{X}^H(f, \tau)]_{P_T^k}\mathbf{W}_k(f)}{\mathbf{W}_k^H(f)E[\mathbf{X}(f, \tau)\mathbf{X}^H(f, \tau)]_{P_N^k}\mathbf{W}_k(f)}, \quad (5.1)$$

kde  $*$  značí konjugaci,  $^H$  konjugaci a transpozici,  $E[\ ]_{P_T^k}$  střední hodnotu zdrojového signálu  $k$  v době, kdy nejsou interference aktivní a  $E[\ ]_{P_N^k}$  představuje střední hodnotu směsi interferujících signálů  $k$ -tého zdroje v době, kdy zdrojový signál  $k$  není aktivní. Zavedením korelační matice pozorovaného zdrojového signálu  $\mathbf{R}_T^k(f)$  a korelační matice pozorované směsi interferujících signálů  $\mathbf{R}_N^k(f)$  se výraz (5.1) zjednoduší na

$$\varepsilon(f) = \frac{\mathbf{W}_k^H(f)\mathbf{R}_T^k(f)\mathbf{W}_k(f)}{\mathbf{W}_k^H(f)\mathbf{R}_N^k(f)\mathbf{W}_k(f)}. \quad (5.2)$$

Matici  $\mathbf{R}_T^k(f)$  resp.  $\mathbf{R}_N^k(f)$  budeme zjednodušeně označovat *korelační matice zdrojového signálu* resp. *korelační matice interferencí*.

Maximální poměr (5.2) získáme derivací  $\varepsilon(f)$  podle  $\mathbf{W}_k^H(f)$  a položením výsledku rovno nule [33]

$$\frac{\mathbf{R}_T^k \mathbf{W}_k (\mathbf{W}_k^H \mathbf{R}_N^k \mathbf{W}_k) - (\mathbf{W}_k^H \mathbf{R}_T^k \mathbf{W}_k) (\mathbf{R}_N^k \mathbf{W}_k)}{(\mathbf{W}_k^H \mathbf{R}_N^k \mathbf{W}_k)^2} = 0, \quad (5.3)$$

kde byla z důvodu přehlednosti vynechána závislost na frekvenci. Jednoduchou úpravou získáme

$$\varepsilon(f) \mathbf{R}_N^k(f) \mathbf{W}_k(f) = \mathbf{R}_T^k(f) \mathbf{W}_k(f), \quad (5.4)$$

$$\varepsilon(f) \mathbf{W}_k(f) = \left[ \left[ \mathbf{R}_N^k(f) \right]^{-1} \mathbf{R}_T^k(f) \right] \mathbf{W}_k(f). \quad (5.5)$$

Vztah (5.5) je shodný s problémem vlastních čísel lineárního systému, kde  $\varepsilon(f)$  je vlastní číslo a  $\left[ \mathbf{R}_N^k(f) \right]^{-1} \mathbf{R}_T^k(f)$  odpovídá vlastnímu vektoru. Pro nalezení maximálního poměru (5.2) vybereme vlastní vektor  $\mathbf{E}_k(f)$  odpovídající největšímu vlastnímu číslu  $\varepsilon(f)$ . Výsledný prostorový filtr určíme podle

$$\mathbf{W}_k(f) = \mathbf{E}_k(f). \quad (5.6)$$

Max SNR tvarovač je navržen pouze na základě vztahu (5.1), který neklade, na rozdíl od tvarovače přijímací charakteristiky typu LCMV, žádnou podmínku na prostorový filtr. To má u širokopásmových signálů za následek, obdobně jako u BSS systému z kapitoly 2.1, problém s vážením užitečného signálu, kdy separované signály nemají stejnou energii jako jejich zdrojové protějšky.

Kompenzaci vážení můžeme provést na základě minimalizace účelové funkce [1]

$$\mathfrak{J}(\mathbf{A}_k(f)) = \mathbb{E} \left[ \left\| \mathbf{X}(f, \tau) - \mathbf{A}_k(f) Y_k(f, \tau) \right\|^2 \right], \quad (5.7)$$

kde  $\| \cdot \|$  značí Euklidovskou vzdálenost,  $\mathbf{A}_k(f) = [A_{k1}(f), \dots, A_{kM}(f)]^T$  je prostorový filtr maximálně přibližující výstupní signál  $Y_k(f, \tau)$  k vektoru pozorovaných signálů

$$\mathbf{X}(f, \tau) \approx \mathbf{A}_k(f) Y_k(f, \tau) = \mathbf{A}_k(f) \mathbf{W}_k^H(f) \mathbf{X}(f, \tau). \quad (5.8)$$

Optimální  $\mathbf{A}_k(f)$  získáme položením derivace účelové funkce (5.7) podle  $\mathbf{A}_k(f)$  rovno nule [1]

$$\mathbf{A}_k(f) \mathbb{E} \left[ Y_k(f, \tau) Y_k^*(f, \tau) \right] - \mathbb{E} \left[ Y_k^*(f, \tau) \mathbf{X}(f, \tau) \right] = 0. \quad (5.9)$$

Vztah dále upravíme do výsledné podoby

$$\mathbf{A}_k(f) = \frac{\mathbb{E} \left[ Y_k^*(f, \tau) \mathbf{X}(f, \tau) \right]}{\mathbb{E} \left[ Y_k(f, \tau) Y_k^*(f, \tau) \right]} = \frac{\mathbf{R}(f) \mathbf{W}_k(f)}{\mathbf{W}_k^H(f) \mathbf{R}(f) \mathbf{W}_k(f)}. \quad (5.10)$$

Z (5.8) je zřejmé, že při kompenzaci vážení získáme  $M$  rozměrný vektor separovaného signálu  $k$ , nebo-li separovaný signál na všech senzorech  $j$ . Naším požadavkům na BSS systém vyhovuje i jednokanálový výstup, a proto vybereme separovaný signál na libovolném vektoru  $J$ .

K návrhu Max SNR tvarovače včetně kompenzace vážení musíme znát korelační matici zdrojového signálu  $\mathbf{R}_T^k(f)$  a korelační matice interferencí  $\mathbf{R}_N^k(f)$ . V případě BSS a nestacionárních interferencí, je určení výše uvedených korelačních matic velmi obtížné a klade vysoké nároky na algoritmy VAD.

## 5.2 Tvarovač přijímací charakteristiky typu LCMV

Tvarovač přijímací charakteristiky typu LCMV (zkráceně LCMV tvarovač) využívá velmi obecného přístupu při návrhu prostorového filtru. Hlavní myšlenkou je zavedení lineární podmínky, která zajistí, aby signál dopadající z daného DOA nebyl tvarovačem ovlivněn. Tato podmínka bývá v anglické literatuře označována jako *linear constraint*.

Dalším požadavkem na tvarovač je maximální potlačení signálů dopadajících z jiných směrů. Za tímto účelem minimalizujeme energii výstupního signálu z tvarovače

$$\mathfrak{I}(\mathbf{W}_k(f)) = E[Y_k(f, \tau)Y_k^*(f, \tau)] = \mathbf{W}_k^H(f)\mathbf{R}(f)\mathbf{W}_k(f), \quad (5.11)$$

kde  $\mathbf{R}(f)$  je korelační matice pozorovaného signálu

$$\mathbf{R}(f) = E[\mathbf{X}(f, \tau)\mathbf{X}^H(f, \tau)]. \quad (5.12)$$

Pouhá minimalizace (5.11) však vede k triviálnímu řešení  $\mathbf{W}_k(f) = [0, \dots, 0]^T$ .

K nalezení netriviálního řešení předpokládejme, že známe směsný vektor  $\mathbf{H}_k(f)$  nebo alespoň jeho aproximaci (2.13). Požadujeme, aby užitečný signál byl propuštěn beze změny

$$\mathbf{W}_k^H(f)\mathbf{H}_k(f) = 1, \quad (5.13)$$

a proto zavedeme lineární podmínku

$$C = \mathbf{W}_k^H(f)\mathbf{H}_k(f) - 1 = 0. \quad (5.14)$$

Minimalizace (5.11) může poté být řešena metodou Lagrangeova multiplikátoru [33] vedoucí k řešení

$$\mathbf{W}_k(f) = \left( \frac{\mathbf{R}^{-1}(f)\mathbf{H}_k(f)}{\mathbf{H}_k^H(f)\mathbf{R}^{-1}(f)\mathbf{H}_k(f)} \right)^*. \quad (5.15)$$

K návrhu LCMV tvarovače tedy postačuje znát pouze směsný vektor  $\mathbf{H}_k(f)$  nebo jeho aproximaci (2.13) či alespoň DOA užitečného signálu [16]. Při slepém návrhu separačního systému však není směsný vektor ani DOA apriori známé a slepý návrh není možný.



Potlačení interferencí LCMV tvarovačem může být zvýšeno, pokud je možné odhadnout časový interval, kdy je užitečný signál  $s_k(t)$  neaktivní. Vyjděme ze vztahu (5.15), kde je využita korelační matice pozorovaného signálu  $\mathbf{R}(f)$ . Pozorovaný signál  $\mathbf{X}(f, \tau)$  je však směs užitečného signálu a interferujících signálů. Zvýšeného potlačení interferujících signálů je možné dosáhnout v případě, kdy korelační matice pozorovaného signálu  $\mathbf{R}(f)$  je nahrazena korelační maticí interferencí  $\mathbf{R}_N^k(f)$ .

### 5.3 Úprava systému MENUET

Systém MENUET využívá časově-frekvenčního binárního maskování k separaci signálů. Procesu maskování, ale také postupů použitých pro stanovení masky, je možné využít k určení veličin, které blíže popisují separační prostředí a mohou být využity pro slepý návrh lineárního prostorového filtru.

#### 5.3.1 Určení korelační matice interferencí a korelační matice zdrojového signálu

Korelační matice interferencí a zdrojového signálu můžeme určit metodou časově-frekvenčního maskování tedy např. systémem MENUET. V případě separace  $N$  signálů bude nutné určit  $N$  korelačních matic. Korelační matici interferencí resp. korelační matici zdrojového signálu určenou systémem MENUET budeme označovat  $\hat{\mathbf{R}}_N^k(f)$  resp.  $\hat{\mathbf{R}}_T^k(f)$ . Symbol  $\hat{\phantom{x}}$  zvýrazňuje skutečnost, že se jedná pouze o odhad korelační matice. Obdobně budeme zvýrazňovat také vektory odhadnuté systémem MENUET.

Maskování u systému MENUET je popsáno vztahem (4.10). Je vybrán pozorovaný signál na libovolném senzoru  $P \in \{1, \dots, M\}$  a provedena multiplikace s maskou  $M_k(f, \tau)$ . Vlastní separace je tedy prováděna pouze jednobokově. Pro určení korelační matice  $\hat{\mathbf{R}}_N^k(f)$  je nutné určit směs interferujících signálů na všech senzorech, a proto je maskování prováděno na všech senzorech

$$\hat{\mathbf{Y}}_k(f, \tau) = M_k(f, \tau) \mathbf{X}(f, \tau), \quad (5.16)$$

kde  $\hat{\mathbf{Y}}_k(f, \tau) = [\hat{Y}_{k1}(f, \tau), \dots, \hat{Y}_{kM}(f, \tau)]^T$  je vektor odhadů zdrojového signálu na  $M$  senzorech, tedy  $\hat{\mathbf{Y}}_k(f, \tau) \approx \mathbf{H}_k(f) S_k(f, \tau)$ .

Směs interferujících signálů  $\hat{N}_{kj}(f, \tau)$  zdrojového signálu  $k$  na senzoru  $j$  lze nyní odhadnout

$$\hat{\mathbf{N}}_k(f, \tau) = \sum_{\bar{b}=1, \bar{b} \neq k}^N \hat{\mathbf{Y}}_{\bar{b}}(f, \tau), \quad (5.17)$$

kde  $\hat{\mathbf{N}}_k(f, \tau) = [\hat{N}_{k1}(f, \tau), \dots, \hat{N}_{kM}(f, \tau)]^T$  je vektor směsí interferujících signálů zdrojového signálu  $k$ . Korelační matici interferencí poté jednoduše určíme z

$$\hat{\mathbf{R}}_N^k(f) = E[\hat{\mathbf{N}}_k(f, \tau) \hat{\mathbf{N}}_k^H(f, \tau)]. \quad (5.18)$$

Korelační matici zdrojového signálu získáme obdobně

$$\hat{\mathbf{R}}_T^k(f) = E[\hat{\mathbf{Y}}_k(f, \tau) \hat{\mathbf{Y}}_k^H(f, \tau)]. \quad (5.19)$$

### 5.3.2 Určení směsných vektorů

Určení směsných vektorů je nezbytné pro návrh LCMV tvarovače. V případě vzájemně řídkých signálů můžeme odhadnout směsný vektor  $\hat{\mathbf{H}}_k(f) = [\hat{H}_{k1}(f), \dots, \hat{H}_{kM}(f)]^T$  pomocí systému MENUET.

Vyjděme ze vztahu (4.8). Vektor středů shluků  $\mathbf{c}_k = [c_{k1}, \dots, c_{kM}]^T$  je vypočten během iterativního procesu algoritmu CA  $k$ -průměrů, jež je určen k stanovení separační masky. Kromě stanovení separační masky je možné střed shluku využít také pro určení směsného vektoru. Dosadíme tedy normalizovaný vektor  $\bar{\mathbf{X}}(f, \tau)$  ze vztahu (4.5) do (4.8)

$$c_{kj} = \left[ \frac{|X_j(f, \tau)|}{\sqrt{\sum_{j=1}^M |X_j(f, \tau)|^2}} \exp \left[ j \frac{\arg(X_j(f, \tau) / X_j(f, \tau))}{2\pi f c^{-1} d_{\max}} \right] \right]_{\bar{X}_j(f, \tau) \in C_k} \quad (5.20)$$

Dále rozepíšeme pozorovaný signál  $X_j(f, \tau)$  dle (2.14)

$$c_{kj} = \left[ \frac{|H_{kj}(f) S_k(f, \tau)|}{\sqrt{\sum_{j=1}^M |H_{kj}(f) S_k(f, \tau)|^2}} \exp \left[ j \frac{\arg\left(\frac{H_{kj}(f) S_k(f, \tau)}{H_{kj}(f) S_k(f, \tau)}\right)}{2\pi f c^{-1} d_{\max}} \right] \right]_{\bar{X}_j(f, \tau) \in C_k} \quad (5.21)$$

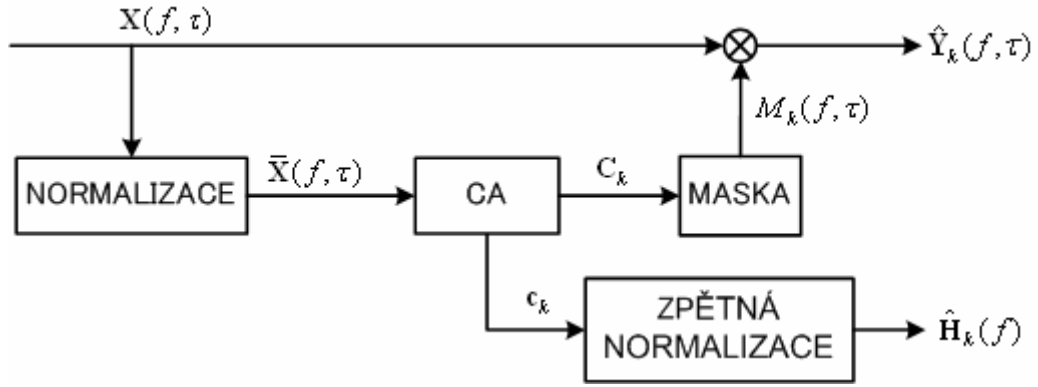
Po vykrácení dominantních zdrojových signálů  $S_k(f, \tau)$  získáme zjednodušený výraz

$$c_{kj} = \left[ \frac{|H_{kj}(f)|}{\sqrt{\sum_{j=1}^M |H_{kj}(f)|^2}} \exp \left[ j \frac{\arg(H_{kj}(f) / H_{kj}(f))}{2\pi f c^{-1} d_{\max}} \right] \right]_{\bar{X}_j(f, \tau) \in C_k}, \quad (5.22)$$

který udává vztah mezi středem shluku  $c_{kj}$  a impulsní odezvou  $H_{kj}(f)$ .

Vektor středů shluků  $\mathbf{c}_k$  představuje odhad směsného vektoru  $\mathbf{H}_k(f)$  ze vztahu (2.15). Směsný vektor  $\mathbf{H}_k(f)$  může být získán z vektoru středu shluků  $\mathbf{c}_k$  inverzním procesem k procesu normalizace (5.22) (zpětnou normalizací)

$$\hat{H}_{kj}(f) = |c_{kj}| \sqrt{\sum_{j=1}^M |c_{kj}|^2} \exp \left[ j 2\pi f c^{-1} d_{\max} \arg(c_{kj} c_{kj}) \right]. \quad (5.23)$$

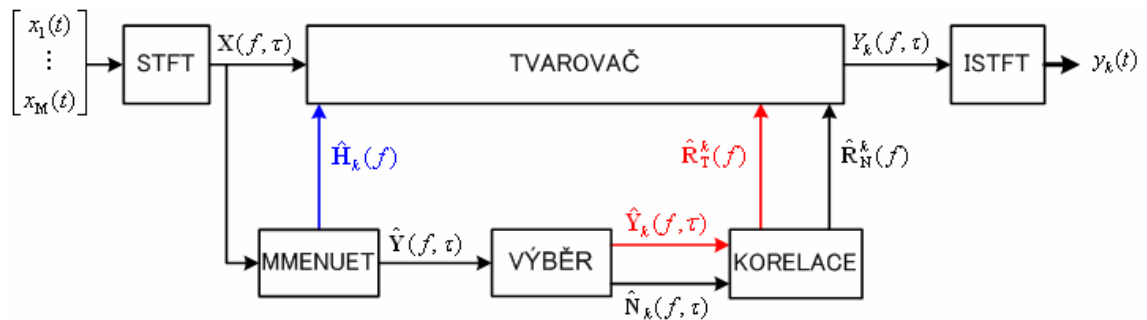


Obr. 5.1. Blokové schéma upraveného systému MENUET pro slepý návrh tvarovačů – MMENUET.

Úpravy systému MENUET představené v kapitole 5.3.1 a kapitole 5.3.2 jsou zobrazeny na blokovém schématu na obr. 5.1. Takto upravený systém budeme nazývat MMENUET. Oproti blokovému schématu na obr. 4.1 byl přidán blok ZPĚTNÁ NORMALIZACE stanovující odhad směšného vektoru  $\hat{\mathbf{H}}_k(f)$  podle (5.23). Dále byl odstraněn blok VÝBĚR, aby mohl být určen vektor odhadů zdrojového signálu  $\hat{\mathbf{Y}}_k(f, \tau)$ .

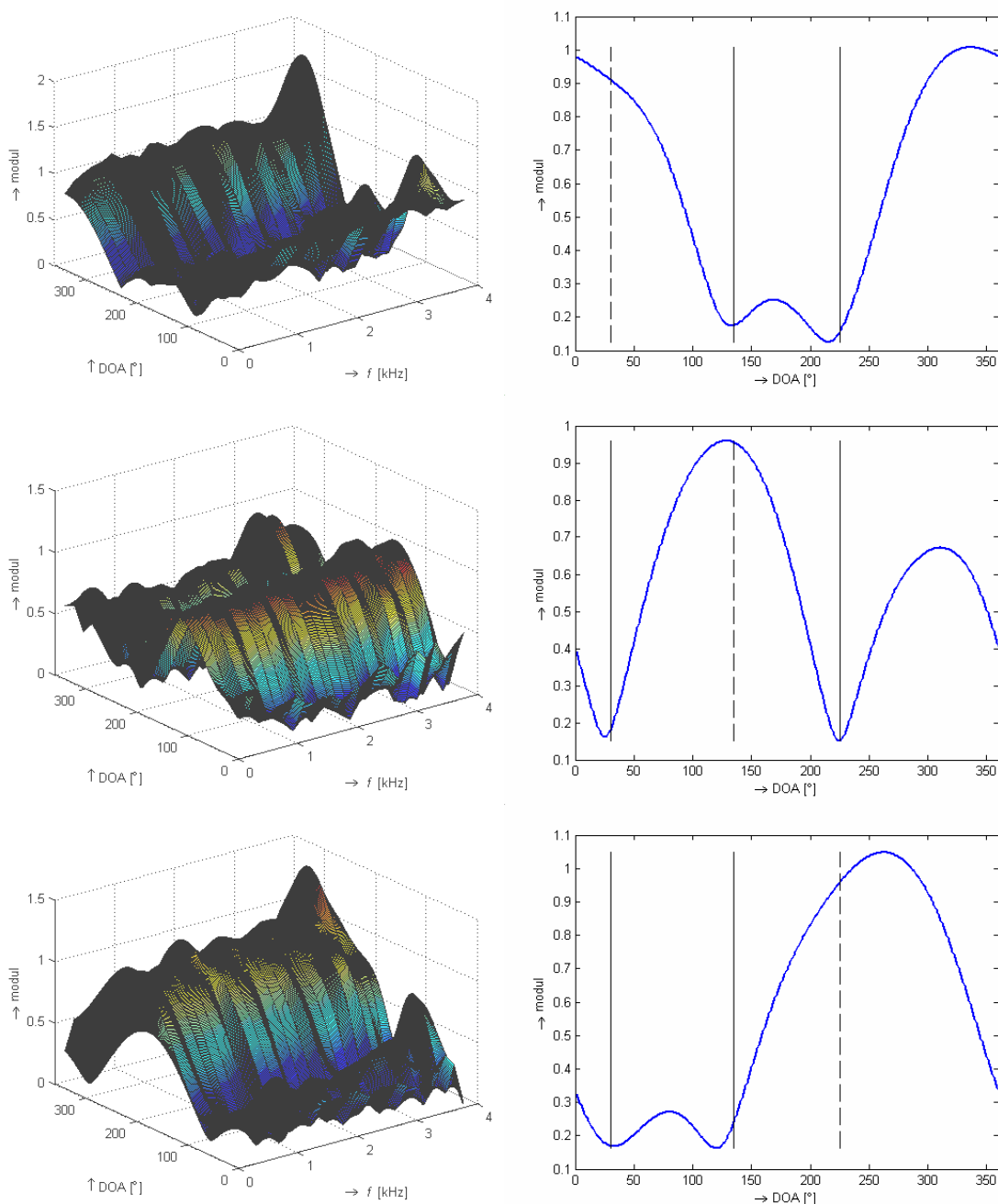
#### 5.4 Sestavení separačního systému

Blokové schéma navrženého BSS systému je uvedeno na obr. 5.2. Jádrem celého systému je tvarovač, jenž zvýrazňuje  $k$ -tý zdrojový signál. Pro přehlednost je na obr. 5.2 zobrazen pouze jeden tvarovač. K separaci  $N$  signálů je však nutně třeba  $N$  tvarovačů. V případě Max SNR resp. LCMV tvarovače je pro návrh využit vztah (5.6) a (5.10) resp. (5.15). Neznámé proměnné z těchto vztahů jsou stanoveny na základě systému MMENUET jehož výstupem je matice odhadů zdrojových signálů  $\hat{\mathbf{Y}}(f, \tau) = [\hat{\mathbf{Y}}_1(f, \tau), \dots, \hat{\mathbf{Y}}_N(f, \tau)]$ . V bloku VÝBĚR je vybrán  $k$ -tý sloupec z matice  $\hat{\mathbf{Y}}(f, \tau)$  a proveden výpočet vektoru směsí interferujících signálů  $\hat{\mathbf{N}}_k(f, \tau)$  dle (5.17). V bloku KORELACE jsou vypočteny korelační matice  $\hat{\mathbf{R}}_N^k(f)$  a  $\hat{\mathbf{R}}_T^k(f)$ .

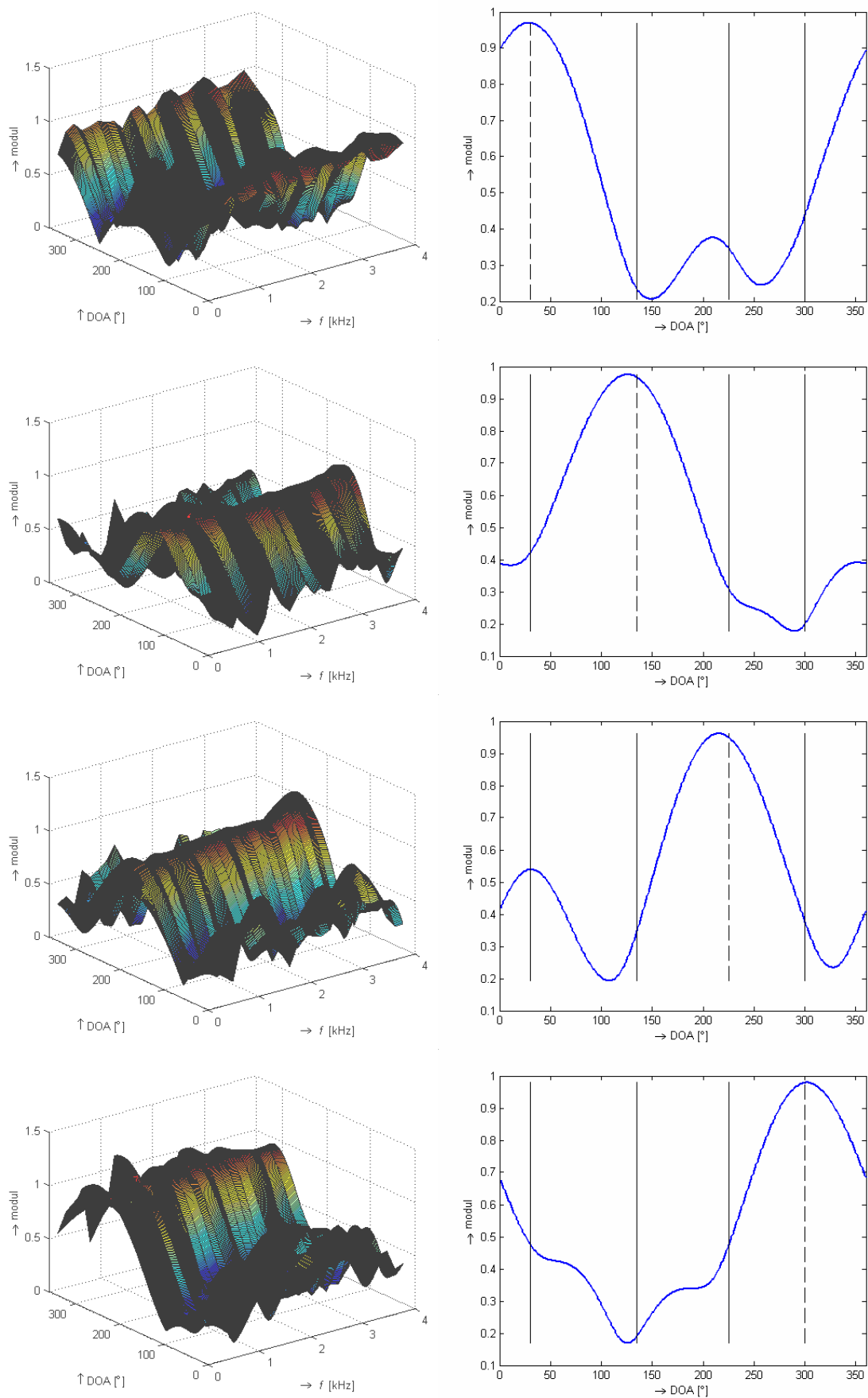


Obr. 5.2. Blokové schéma navrženého BSS systému kombinujícího tvarovače a MMENUET. Modrá barva resp. červená barva představuje část blokového schématu, která je využita pouze pro LCMV resp. Max SNR tvarovač.

Prostorové filtry získané navrženým BSS systémem jsou zobrazeny na obr. 5.3 resp. obr. 5.4 pro DC resp. UDC uspořádání. Prostorové filtry byly navrženy pro totožný experiment, jenž je uvedený v kapitole 4.3. Černou čarou je zobrazeno DOA interferencí signálů. Černou přerušovanou čarou je zobrazeno DOA zdrojového signálu.



Obr. 5.3. Navržené prostorové filtry BSS systému kombinujícího LCMV tvarovače a MMENUET, DC uspořádání. Vlevo: 3D zobrazení. Vpravo: průměrná hodnota přes všechny frekvenční složky. Černá – DOA interferencí, černá přerušovaná – DOA zdrojového signálu.



Obr. 5.4. Navržené prostorové filtry. Obdoba obr. 5.3 pro UDC uspořádání.

## 5.5 Experimentální část

V této části vyhodnotíme představený BSS systém pomocí hodnot SIR a SDR. Experimenty byly provedeny za totožných podmínek jako v kapitole 4.3. Systémové uspořádání je uvedeno na obr. 4.9. Výsledky experimentů jsou shrnuty v tab. 5.1. Zkratka LCMV označuje separační systém, kdy pro návrh LCMV tvarovačů nebyl využit MMENUET. Návrh byl proveden na základě znalosti DOA zdrojových signálů a korelační matice pozorovaných signálů  $\mathbf{H}_k(f)$ . Návrh tedy *nebyl proveden slepě*. M-LCMV resp. M-Max SNR je označení BSS systému kombinujícího LCMV resp. Max SNR tvarovače a MMENUET. Index Ideal ukazuje výkon BSS systému v případě, kdy jsou proměnné potřebné k návrhu BSS systému (korelační matice interferencí, korelační matice zdrojového signálu, směsný vektor) odhadnuty zcela přesně.

Tab. 5.1. Výsledky experimentů.

	DC				UDC			
	SIR [dB]		SDR [dB]		SIR [dB]		SDR [dB]	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
LCMV	8,3	0,8	6,8	1,0	7,9	0,5	7,3	9,9
M-LCMV	18,3	2,0	16,2	0,8	12,7	0,6	13,4	0,6
M-LCMV <sub>Ideal</sub>	23,8	0,7	16,5	2,0	13,7	1,1	15,1	1,7
M-Max SNR	20,8	1,7	17,1	2,1	13,0	1,0	15,5	0,6
M-Max SNR <sub>Ideal</sub>	23,9	0,7	17,6	1,9	13,7	1,0	16,6	1,6

Z tab. 5.1 je zřejmé, že BSS systémy M-LCMV i M-Max SNR jsou výrazně lepší v DC i UDC uspořádání než LCMV. M-Max SNR dosahuje hlavně v DC uspořádání lepších výsledků než M-LCMV. V ideálním případě, kdy návrh však není proveden slepě, jsou systémy z hlediska SIR a SDR přibližně shodné.

## 5.6 Souhrn

V této kapitole byl představen BSS systém kombinující MMENUET a LCMV či Max SNR tvarovače. Separace tímto systémem může být provedena dokonce i v UDC uspořádání. Separované signály neobsahují žádný hudební šum, neboť separační proces je lineární. Provedené experimenty ukazují, že kombinace MMENUETU a tvarovačů výkonnostně převyšují LCMV beamformer, který navíc není navržen slepě. Dále bylo

zjištěno, že M-Max SNR výkonnostně převyšuje M-LCMV, a to hlavně v DC uspořádání. V případě ideálního návrhu systému se však výkonnostně systémy přibližně shodují.

Pokud porovnáme navržené systémy se systémem MENUET, zjistíme, že v UDC uspořádání dosahují nižšího potlačení interferencí. Byl odstraněn hudební šum, ale za cenu zvýšeného pronikání interferujících signálů do separovaného signálu, což není samozřejmě ideální. Otázkou, jak potlačit hudební šum a zároveň zamezit zvýšenému pronikání interferujících signálů se bude zabývat další kapitola.

## 6 Kaskádní spojení systému MENUET a tvarovačů přijímací charakteristiky

V kapitole 5 byl představen BSS systém, který může být použit dokonce i v UDC uspořádání a vlastní separace signálů je prováděna lineární prostorovou filtrací. Separované signály nejsou tedy degradovány hudebním šumem, což je velká výhoda oproti BSS systémům založeným na systému MENUET, viz kapitola 4. Pokud však porovnáme hodnoty SIR v tab. 4.1 a tab. 5.1, zjistíme, že BSS systém v kapitole 5 dosahuje v UDC uspořádání horších výsledků než samotný MENUET. Interference od ostatních zdrojů jsou tedy výraznější.

Důvodem pro nižší potlačení interferujících signálů je právě UDC uspořádání. Při návrhu adaptivního tvarovače na  $M$  senzorovém poli je možné navrhnout maximálně  $M$  extrémních bodů, z čehož jeden je právě určen pro extrahování zdrojového signálu  $k$ . Ostatních  $M - 1$  extrémních bodů (nulových bodů) může být využito na potlačení interferujících signálů. V případě UDC uspořádání, kdy  $N > M$ , převyšuje počet interferujících zdrojů počet nulových bodů. Nulové body nejsou proto umístěny v odhadnutých DOA interferujících signálů, jak je tomu v DC uspořádání, ale jsou interpolovány tak, aby došlo k statisticky optimálnímu potlačení interferencí, viz obr. 5.4.

V této kapitole je představeno řešení, jakým lze odstranit výše uvedené nevýhody a současně zachovat lineární proces separace signálů.

### 6.1 Převod na určenou BSS

Vyjděme ze systému MMENUET, pomocí kterého je možné určit matici odhadů zdrojových signálů  $\hat{\mathbf{Y}}(f, \tau)$  [44], [46]. Jedná se vlastně již o separované signály, které můžeme opět libovolně smíchat pouhou sumací jako např. ve vztahu (5.17). Proto se nabízí možnost kaskádního spojení systému MMENUET a tvarovače k převedení UDC na DC uspořádání.

Převod na DC uspořádání je možné provést omezením počtu interferujících signálů podle

$$\hat{\mathbf{N}}_k(f, \tau) = \sum_{\bar{b} \in z_k} \hat{\mathbf{Y}}_{\bar{b}}(f, \tau), \quad (6.1)$$

kde  $z_k = \{q | 1 \dots N, q \neq k\}^{M-1}$  je množina s  $M - 1$  indexy interferujících signálů, které budou potlačeny tvarovačem, jež extrahuje  $k$ -tý zdrojový signál. Indexy interferujících signálů jsou vybrány na základě počtu členů v shlucích  $C_1, \dots, C_N$ .

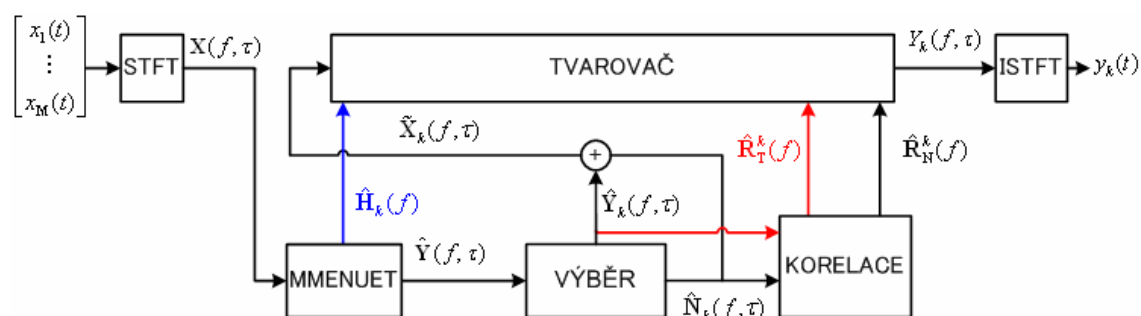
K převedení UDC na DC uspořádání je nutné nezahrnout do množiny  $z_k$   $N - M$  interferujících signálů. Tato redukce je podmíněna vkládáním nulových hodnot



do spektra signálu  $\hat{N}_k(f, \tau)$  což ovšem může vést k degradaci hudebním šumem. Pro omezení hudebního šumu je proto výhodné zahrnovat do množiny  $z_k$  takové interferující signály, jejichž shluky obsahují nejvyšší počet členů. V porovnání se systémem MENUET je však nulováno méně časově-frekvenčních složek, a tudíž hudební šum by neměl být tak znatelný. Pro úplnost uvedme, že v případě systému MENUET je nulováno všech  $N - 1$  interferujících signálů.

Blokové schéma kaskádního spojení systému MMENUET a tvarovače je uvedeno na obr. 6.1. Vstupní signál tvarovače není tvořen vektorem pozorovaných signálů  $\mathbf{X}(f, \tau)$  jako je tomu v kapitole 5, ale je určen podle

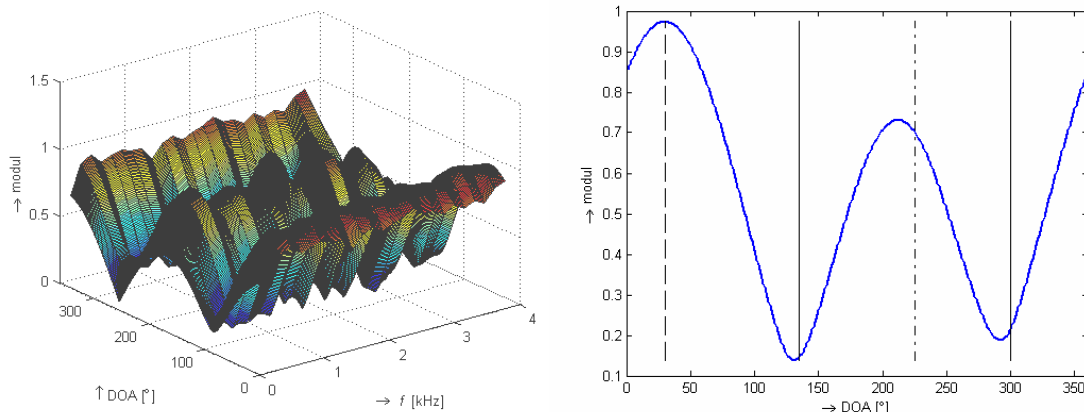
$$\tilde{\mathbf{X}}_k(f, \tau) = \hat{\mathbf{Y}}_k(f, \tau) + \hat{\mathbf{N}}_k(f, \tau). \quad (6.2)$$

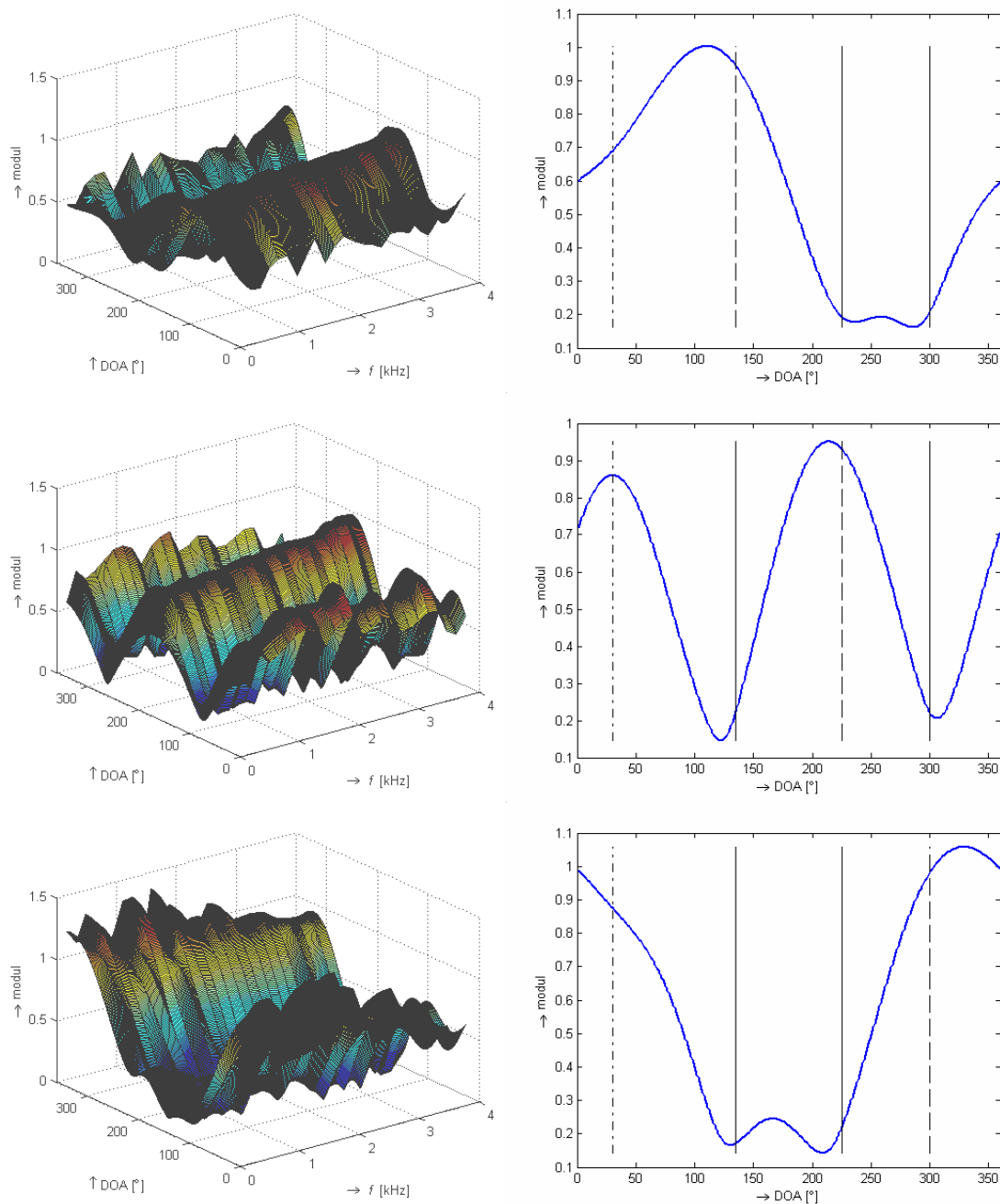


Obr. 6.1. Blokové schéma systému pro nedourčenou BSS.

Prostorové filtry získané navrženým BSS systémem jsou zobrazeny na obr. 6.2. Návrh byl proveden pro UDC uspořádání dle obr. 4.9. Experiment proběhl za totožných podmínek jako experimenty uvedené v kapitole 4.3. Na obr. 6.2 je zobrazeno černou čarou DOA interferujících signálů. Černou čárkovanou čarou je zobrazeno DOA zdrojového signálu a černou čerchovanou čarou DOA signálu jež nebyl zahrnut do vstupního signálu  $\tilde{\mathbf{X}}_k(f, \tau)$ .

Z obr. 6.2 je zřejmé, že nulové body jsou navrženy v DOA interferujících signálů, jež jsou zahrnuty ve vstupním signálu  $\tilde{\mathbf{X}}_k(f, \tau)$ . Dochází tedy k efektivnějšímu potlačení interferencí, než je tomu u systému v kapitole 5.4.





Obr. 6.2. Navržené prostorové filtry BSS systém kombinujícího MMENUET a LCMV tvarovače v kaskádním zapojení, UDC uspořádání. Vlevo: 3D zobrazení. Vpravo: průměrná hodnota přes všechny frekvenční složky. Černá – DOA interferencí, černá přerušovaná – DOA zdrojového signálu, černá čerchovaná – DOA interferencí odstraněných systémem MMENUET.

## 6.2 Pole tvarovačů přijímací charakteristiky

Doposud představené BSS systémy kombinující systém MMENUET a tvarovače používaly vždy pouze jeden tvarovač pro zvýraznění zdrojového signálu  $k$ . V této

kapitole představíme BSS systém, který využívá pro extrahování jednoho zdrojového signálu  $D$  LCMV tvarovačů. Tento BSS systém budeme označovat jako *pole tvarovačů* (BA) [47].

BA je zobrazeno na obr. 6.3. Skládá se z  $N$  sloupců a  $D$  řádek. Každý ze sloupců slouží k extrahování jednoho zdrojového signálu  $k$  pomocí  $D$  LCMV tvarovačů.  $D$  je kombinační číslo [9] určené podle

$$D = \binom{N-1}{Z} = \frac{(N-1)!}{Z!(N-1-Z)!}, \quad (6.3)$$

kde  $Z \in \{1, \dots, M-1\}$  je počet interferujících signálů obsažených ve vektoru vstupních signálů  $\tilde{\mathbf{X}}_{kd}(f, \tau)$ . Každý prvek BA je navržen podle

$$\mathbf{W}_{kd}(f) = \left( \frac{[\hat{\mathbf{R}}_N^{kd}(f)]^{-1} \hat{\mathbf{H}}_k(f)}{\hat{\mathbf{H}}_k^H(f) [\hat{\mathbf{R}}_N^{kd}(f)]^{-1} \hat{\mathbf{H}}_k(f)} \right)^*, \quad (6.4)$$

kde  $d$  je index označující řádku v BA a  $k$  je index sloupce.

Vstupní signál každého pole BA je

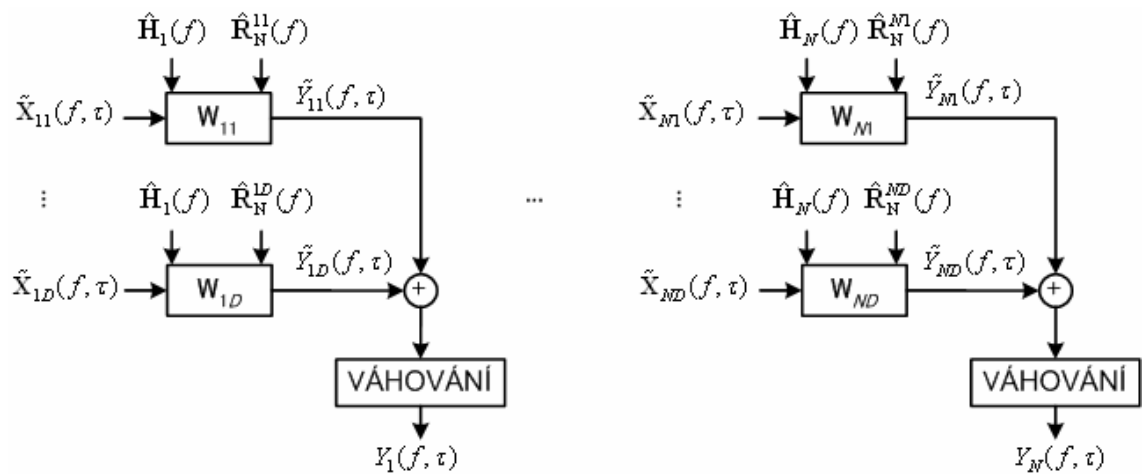
$$\tilde{\mathbf{X}}_{kd}(f, \tau) = \hat{\mathbf{Y}}_k(f, \tau) + \hat{\mathbf{N}}_{kd}(f, \tau), \quad (6.5)$$

kde  $\hat{\mathbf{N}}_{kd}(f, \tau)$  je vektor směsí interferujících signálů. V každé směsi je obsaženo  $Z$  interferujících signálů.  $\hat{\mathbf{N}}_{kd}(f, \tau)$  získáme

$$\hat{\mathbf{N}}_{kd}(f, \tau) = \sum_{b \in Z_{kd}} \hat{\mathbf{Y}}_b(f, \tau), \quad (6.6)$$

kde  $Z_{kd}$  je  $d$ -tá  $Z$ -prvková podmnožina [9] z  $N-1$  prvkové množiny  $\theta_k = \{q | 1 \dots N, q \neq k\}$ . Množina  $\theta_k$  obsahuje indexy všech interferujících signálů. Korelační matici interferencí pro  $d$ -tý řádek a  $k$ -tý sloupec určíme podle

$$\hat{\mathbf{R}}_N^{kd}(f) = E[\hat{\mathbf{N}}_{kd}(f, \tau) \hat{\mathbf{N}}_{kd}^H(f, \tau)]. \quad (6.7)$$



Obr. 6.3. Pole tvarovačů.

Uvedme si příklad pro  $N = 4, M = 3, Z = 1$  a  $k = 4$ . BA má tedy  $N = 4$  sloupců a  $D = 3$  řádků. Množina indexů interferujících signálů je  $\theta_k = \{1, 2, 3\}$  a 1-podmnožiny jsou  $z_{11} = 1, z_{12} = 2, z_{13} = 3$ . Vstupní signály LCMV tvarovačů jsou:

- $\tilde{\mathbf{X}}_{41}(f, \tau) = \hat{\mathbf{Y}}_4(f, \tau) + \hat{\mathbf{Y}}_1(f, \tau),$
- $\tilde{\mathbf{X}}_{42}(f, \tau) = \hat{\mathbf{Y}}_4(f, \tau) + \hat{\mathbf{Y}}_2(f, \tau),$
- $\tilde{\mathbf{X}}_{43}(f, \tau) = \hat{\mathbf{Y}}_4(f, \tau) + \hat{\mathbf{Y}}_3(f, \tau).$

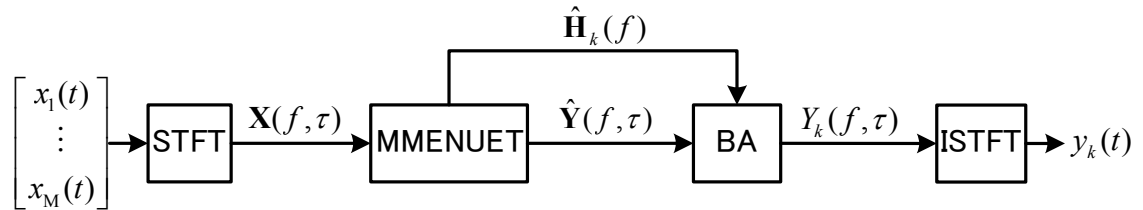
V každém sloupci  $k$  jsou na obr. 6.3 výstupní signály LCMV tvarovačů  $\tilde{Y}_{kd}(f, \tau) = \mathbf{W}_{kd}^T(f) \tilde{\mathbf{X}}_{kd}(f, \tau)$  sečteny a váženy počtem řad  $D$

$$Y_k(f, \tau) = \frac{\sum_{d=1}^D \tilde{Y}_{kd}(f, \tau)}{D}, \quad (6.8)$$

kde  $Y_k(f, \tau)$  je výstupní separovaný signál. Vážení je na obr. 6.3 provedeno v bloku VÁHOVÁNÍ. Výstupní signály z LCMV tvarovačů mohou být sečteny a posléze váženy neboť platí podmínka (5.13).

Vztah (6.8) je důležitý ze dvou důvodů. V prvním případě dojde k dalšímu zvýraznění zdrojového signálu  $k$ . Výstup LCMV tvarovačů  $\tilde{Y}_{kd}(f, \tau)$  již sám o sobě představuje extrahovaný zdrojový signál  $k$  se zbytky interferencí pocházejících z vektoru směsi interferujících signálů  $\hat{\mathbf{N}}_{kd}(f, \tau)$ . V každém vektoru vstupního signálu LCMV tvarovačů  $k$ -tého sloupce  $\tilde{\mathbf{X}}_{kd}(f, \tau)$  je obsažen vektor odhadu zdrojového signálu  $k$   $\hat{\mathbf{Y}}_k(f, \tau)$  a rozdílné vektory směsi interferujících signálů  $\hat{\mathbf{N}}_{kd}(f, \tau)$ . Sečtením signálů  $\tilde{Y}_{kd}(f, \tau)$  tedy opět dojde k zvýraznění extrahovaného signálu.

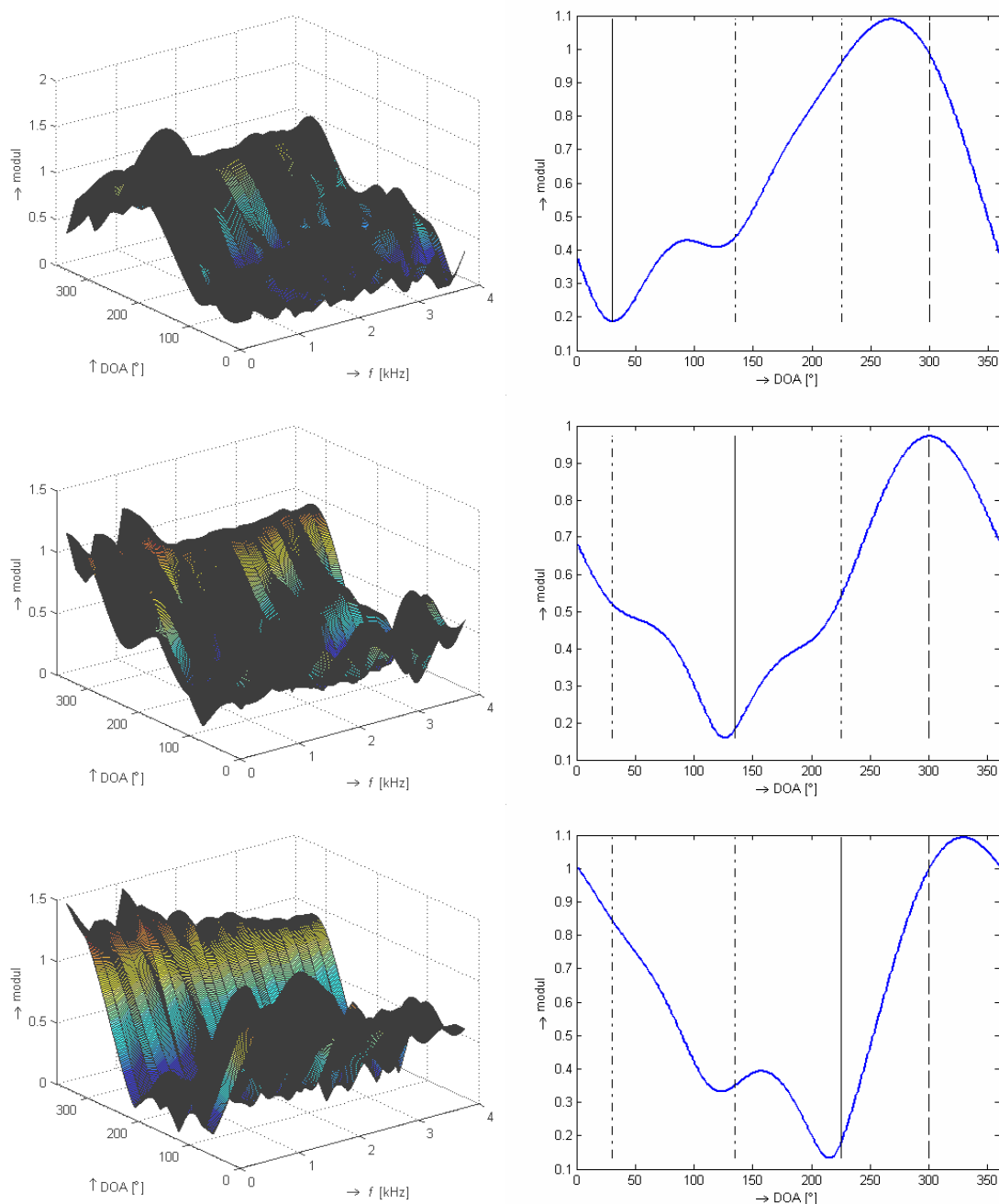
V druhém případě se jedná o odstranění hudebního šumu i pro UDC uspořádání. Při tvorbě vektorů směsi interferujících signálů dle (6.6) je zajištěno, že v každém sloupci  $k$  jsou obsaženy všechny interferující signály. Ve výstupních signálech LCMV tvarovačů  $\tilde{Y}_{kd}(f, \tau)$  je tedy hudební šum obsažen. Po jejich sumaci (6.8) je ale hudební šum odstraněn. Blokové schéma výsledného BSS systému využívajícího BA je zobrazeno na obr. 6.4.



Obr. 6.4. Blokové schéma BSS systému založeného na BA.

Prostorové filtry  $k$ -tého sloupce BA navržené BSS systémem z obr. 6.4 jsou uvedeny na obr. 6.5. Návrh byl proveden pro UDC uspořádání,  $N = 4, M = 3, Z = 1$  a  $k = 4$ . Experiment proběhl za totožných podmínek jako

experimenty uvedené v kapitole 4.3. Systémové uspořádání je na obr. 4.9. DOA interferujících signálů je na obr. 6.5 zobrazeno černou čarou. Černou čárkovanou čarou je zobrazeno DOA zdrojového signálu a černou čerchovanou čarou pak DOA signálů, které nebyly zahrnuty do vstupního signálu  $\tilde{\mathbf{X}}_{kd}(f, \tau)$ .



Obr. 6.5. Navržené prostorové filtry pro  $k$ -tý sloupec BA.  $N = 4$ ,  $M = 3$ ,  $Z = 1$  a  $k = 4$ . Vlevo: 3D zobrazení. Vpravo: průměrná hodnota přes všechny frekvenční složky. Černá – DOA interferencí, černá přerušovaná – DOA zdrojového signálu, černá čerchovaná – DOA interferencí odstraněných systémem MMENUET.

### 6.3 Experimentální část

V této části vyhodnotíme představené BSS systémy pomocí hodnot SIR a SDR. Experimenty byly provedeny za totožných podmínek jako v kapitole 4.3. Systémové uspořádání je uvedeno na obr. 4.9. Výsledky experimentů s BSS systém z kapitoly 6.1 jsou uvedeny v tab. 6.1, kde M-Max SNR dosahuje obdobně jako v kapitole 5.5 o něco vyšší SIR i SDR. Výsledky pro DC uspořádání jsou totožné s výsledky uvedenými v tab. 5.1.

Tab. 6.1. Výsledky experimentů BSS systému převádějícího UDC na DC uspořádání.

	UDC			
	SIR [dB]		SDR [dB]	
	$\mu$	$\sigma$	$\mu$	$\sigma$
M-LCMV	15,6	0,7	11,8	0,5
M-Max SNR	16,0	1,1	13,0	0,6

Vyhodnocení BSS systému využívajícího BA je uvedeno v tab. 6.2. V tab. 6.2 používáme značení BAZ, kde  $Z$  označuje  $Z$ -prvkovou podmnožinu. Každý prvek BA1 resp. BA2 potlačuje jeden resp. dva interferující signály. V DC uspořádání se skládá BA1 resp. BA2 z tří sloupců a dvou resp. jednoho řádku. V UDC uspořádání je rozměr BA1 i BA2 shodný tj. čtyři sloupce a tři řádky. Z výsledků v tab. 6.2 je zřejmé, že zvýšením počtu interferujících signálů ve vektoru směsi interferujících signálů  $\hat{N}_{kd}(f, \tau)$  dochází k snížení SIR a k zvýšení SDR.

Tab. 6.2. Výsledky experimentů BSS systému založeného na BA.

	DC				UDC			
	SIR [dB]		SDR [dB]		SIR [dB]		SDR [dB]	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
BA1	20,0	2,0	12,1	0,1	20,5	0,9	10,7	0,5
BA2	18,3	2,1	16,2	0,8	17,7	0,8	12,2	0,7

## 6.4 Souhrn

V kapitole 6.1 jsme představili BSS systém převádějící UDC uspořádání na DC uspořádání. Převod je proveden odstraněním  $N - M$  interferujících signálů ze vstupního signálu  $k$ -tého tvarovače. Zbylé interferující signály jsou potlačeny samotným tvarovačem. Tento systém dosahuje vyšší SIR, avšak nižší SDR než BSS systém v kapitole 5. Při převodu UDC uspořádání na DC uspořádání dochází k vkládání nulových hodnot do spektra vektoru pozorovaných signálů, což může vést k degradaci separovaného signálu hudebním šumem.

V kapitole 6.2 byl představen BSS systém založený na BA. Hodnoty SIR i SDR dosahují při porovnání s ostatními představenými systémy velmi dobrých výsledků. Tento systém navíc nevnáší do separovaného signálu hudební šum, a to dokonce ani v UDC uspořádání. Cenou za nárůst separačního výkonu a odstranění hudebního šumu jsou zvýšené výpočetní nároky, neboť pro separaci každého z  $N$  signálů je třeba  $D$  tvarovačů.

## 7 Jednokanálové potlačení interferencí v separovaném signálu

Představené vícekanálové separační techniky využívají prostorové filtrace k oddělení interferujících signálů od užitečného signálu. Rušení přicházející ze stejného směru jako užitečný signál není potlačeno. Pro potlačení tohoto rušení je možné využít jednokanálové metody jako např. [11], [26], [41].

V této kapitole popíšeme jednoduchou a velice rychlou jednokanálovou techniku, která u systémů představených v kapitole 5 a 6 dále potlačuje interference přicházející z jiného směru než je DOA užitečného signálu. Jedná se o obdobu časově-frekvenčního maskování, při kterém jsou interference potlačeny v časově-frekvenčních složkách kde  $M_k(f, \tau) = 0$ . V BSS systémech z kapitoly 5 a 6 je separační maska  $M_k(f, \tau)$  známa, a proto můžeme stanovit novou masku

$$\xi_k(f, \tau) = \begin{cases} 1 & M_k(f, \tau) = 1 \\ \beta & M_k(f, \tau) = 0 \end{cases}, \quad (7.1)$$

kde  $0 \leq \beta \leq 1$ . Zvýrazněný užitečný signál získáme

$$Y_E^k(f, \tau) = \xi_k(f, \tau) Y_k(f, \tau). \quad (7.2)$$

Při nízké hodnotě koeficientu  $\beta$  však můžeme zavedením masky  $\xi_k(f, \tau)$  opět vnést do separovaného signálu hudební šum, což je samozřejmě nežádoucí.

### 7.1 Experimentální část

BSS systémy z kapitol 5 a 6 doplníme o jednokanálové zvýraznění separovaného signálu pomocí masky  $\xi_k(f, \tau)$ . Výsledky budeme prezentovat pomocí hodnot SIR a SDR. Experimenty byly provedeny za totožných podmínek jako v kapitole 4.3. Systémové uspořádání je uvedeno na obr. 4.9. Výsledky experimentů pro  $\beta = 0,5$  a  $\beta = 0$  jsou uvedeny v tab. 7.1. Pokud je  $\beta = 1$  není jednokanálové potlačení interferencí aplikováno a výsledky experimentů jsou shodné s výsledky v tab. 5.1 a v tab. 6.2.

Z tab. 7.1 je zřejmé, že snižováním koeficientu  $\beta$  dochází k zvyšování SIR. Zároveň však můžeme pozorovat také pokles SDR. Nastavením  $\beta = 0$  jsou separované signály zvýrazněny binární maskou, obdobně jako je tomu např. u BSS systému MENUET. Na rozdíl od systému MENUET však BSS systémy z tab. 7.1 potlačují interference i v časově-frekvenčních složkách, kde  $M_k(f, \tau) = 1$ . Proto tyto BSS systémy dosahují o mnoho vyšší SIR než samotný MENUET viz tab. 4.1.



Tab. 7.1. Výsledky experimentu s jednokanálovým potlačením interferencí v separovaných signálech.

	$\beta$	DC				UDC			
		SIR [dB]		SDR [dB]		SIR [dB]		SDR [dB]	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
M-LCMV	0,5	22,2	2,4	13,6	1,1	17,0	0,6	11,7	0,6
	0	26,5	2,9	10,4	1,3	23,2	0,7	8,7	0,7
M-Max SNR	0,5	24,6	2,0	14,5	1,4	17,3	1,0	13,1	0,4
	0	28,8	2,7	10,7	1,3	23,6	1,3	9,2	0,7
BA1	0,5	21,6	2,1	10,8	0,4	21,6	1,0	9,7	0,6
	0	22,5	2,2	9,6	0,9	22,3	1,0	8,6	0,6
BA2	0,5	22,2	2,4	13,6	1,1	20,7	0,8	10,5	0,5
	0	26,5	2,9	10,4	1,3	23,5	0,9	8,7	0,5

Výsledky systému M-LCMV jsou v DC uspořádání shodné s výsledky systému BA2. V případě DC uspořádání je i u systému BA2 zdrojový signál zvýrazňován pouze jedním LCMV tvarovačem ( $D=1$ ), a proto jsou systémy M-LCMV a BA2 totožné. V UDC uspořádání je dosaženo silného potlačení interferencí až při  $\beta = 0$ . Systém M-Max SNR dosahuje dobrých výsledků v DC uspořádání. V UDC uspořádání je dosaženo vysokého SIR, obdobně jako u M-LCMV, až při  $\beta = 0$ .

BA systémy dosahují dobrých výsledků jak v DC tak UDC uspořádání. Z tab. 7.1 si můžeme také všimnout, že při  $\beta = 0$  BA2 dosahuje lepších výsledků než BA1. Signály separované systémem BA2 obsahují tedy méně interferencí v časově-frekvenčních složkách kde  $M_k(f, \tau) = 1$ . Dále si můžeme také povšimnout, že výkon BA1 systému je méně ovlivňován koeficientem  $\beta$ . BA systém s  $Z = 1$  tedy lépe potlačuje interference v časově-frekvenčních složkách, kde  $M_k(f, \tau) = 0$  než BA systém s  $Z = M - 1$ . Naproti tomu BA systém s  $Z = M - 1$  potlačuje lépe interference v časově-frekvenčních složkách, kde  $M_k(f, \tau) = 0$ .

## 7.2 Souhrn

V této kapitole bylo představeno jednokanálové zvýraznění separovaného signálu, jež bylo inspirováno znalostí separační masky  $M_k(f, \tau)$  určené systémem MMENUET. Jak bylo ukázáno v experimentální části, interference v separovaném

signálu jsou při nízkých hodnotách koeficientu  $\beta$  silně potlačeny. Současně nárůst výpočetní náročnosti je minimální. Při nastavení koeficientu  $\beta = 0$  dostáváme separační systém na bázi binární masky, který navíc potlačuje interference i v časově-frekvenčních složkách náležejících do shluku užitečného signálu.

Jednokanálové zvýraznění signálu maskou  $\xi_k(f, \tau)$  však představuje nelineární změny ve spektru, které mohou při volbě nízkého koeficientu  $\beta$  způsobit zanesení hudební šumu do separovaného signálu. Koeficient  $\beta$  by proto měl být stanoven tak, aby maximálně potlačoval interference a zároveň nedegradoval hudebním šumem. Určení takové hodnoty však závisí na více faktorech jako je např. použitý BSS systém, separační prostředí, užitečný signál. Určený koeficientu  $\beta$  je proto vhodné dělat experimentálně pro daný BSS problém.

Jinou možností je určit pro každou časově-frekvenční složku, pro kterou platí  $M_k(f, \tau) = 0$ , vlastní koeficient  $\beta(f, \tau)$  v závislosti na energii okolních časově-frekvenčních složek náležejících do shluku užitečného signálu. Tento postup je však výpočetně o mnoho náročnější.

## 8 Srovnání navržených BSS systémů

Vyhodnocení navržených BSS systémů [50] je prováděno pomocí objektivních a subjektivních měr [12], [38]. Objektivní míry jsou vyhodnocovány výpočetními prostředky na základě porovnávání zpracovaného signálu s originálním či vyhodnocením pouze zpracovaného signálu. Subjektivní míry jsou výsledkem subjektivních testů. Jedná se o hodnocení prováděné skupinou posluchačů, kteří známkuji zpracovanou řeč podle stanovené stupnice. Subjektivní testy mohou být prováděny na samotném zpracovaném signálu nebo na základě porovnání s referenčním signálem.

V předchozích kapitolách jsme vyhodnocovali BSS systémy pomocí objektivních měr SIR a SDR udávajících míru potlačení interferujících signálů a degradaci užitečného signálu separačním systémem. Vyhodnocení pouze pomocí SIR a SDR však není zcela dostačující neboť, tyto míry nepostihují mimo jiné hladinu hudebního šumu v separovaném signálu a kvalitu separovaného signálu tak, jak je vnímána posluchačem.

Vyhodnocení hladiny hudebního šumu a kvality separovaného signálu z hlediska posluchače může být provedeno subjektivními testy. Tyto testy jsou však při velkém počtu vyhodnocovaných systémů časově náročné. Proto bylo přistoupeno k redukci počtu systémů na základě ITU-T doporučení P.862 [25].

Doporučení P.862 popisuje objektivní metodu PESQ jež predikuje subjektivní hodnocení posluchačů. Vyhodnocení je prováděno na základě porovnání originálního signálu se zpracovaným signálem dle doporučeného algoritmu [25]. PESQ se používá hlavně pro vyhodnocování úzkopásmových řečových kodeků či signálů, jež prošly přenosovým telekomunikačním médiem. Výstupem je známka v rozsahu 0 až 5 obdobně jako je tomu u parametru MOS [32]. Čím je známka vyšší tím lepší je predikované subjektivní hodnocení.

Výsledky hodnocení BSS systémů metodou PESQ jsou uvedeny v tab. 8.1. Experimenty byly provedeny za totožných podmínek jako v kapitole 4.3. Rozdíl je pouze v souboru hodnot PESQ, který se skládá z  $N$  zdrojových signálů a sice mužských resp. ženských promluv v případě DC resp. UDC uspořádání. Průměrná hodnota resp. směrodatná odchylka z  $N$  promluv je značena jako  $\mu$  resp.  $\sigma$ .

Jak můžeme vidět z tab. 8.1, metoda SMMENUET získala nižší hodnocení než MENUET. FMMENUET naproti tomu dosahuje lepších výsledků. M-LCMV dosahuje totožných výsledků s BA2 v DC uspořádání. Tato shoda je patrná ze struktury obou systémů. M-Max SNR má v DC uspořádání o trochu vyšší hodnocení než M-LCMV. V UDC uspořádání jsou oba systémy prakticky shodné. Systémy založené na BA dosahují

lepších výsledků hlavně v UDC uspořádání. V DC uspořádání se hodnocení o mnoho neliší.

Tab. 8.1. Vyhodnocení metodou PESQ.

	$\beta$	DC		UDC	
		$\mu$	$\sigma$	$\mu$	$\sigma$
MENUET	-	2,92	0,15	2,52	0,26
SMMENUET (22;0,3)	-	2,71	0,14	2,47	0,19
FMMENUET (1,5)	-	2,98	0,15	2,63	0,25
M-LCMV	1	3,03	0,18	2,19	0,24
	0,5	3,18	0,18	2,46	0,24
	0	3,10	0,16	2,67	0,24
M-Max SNR	1	3,07	0,15	2,20	0,24
	0,5	3,20	0,15	2,47	0,24
	0	3,08	0,12	2,66	0,24
BA1	1	3,12	0,18	2,76	0,22
	0,5	3,17	0,19	2,76	0,21
	0	3,06	0,17	2,65	0,24
BA2	1	3,03	0,18	2,62	0,23
	0,5	3,18	0,18	2,75	0,21
	0	3,10	0,16	2,67	0,24

Koeficient  $\beta$  potlačuje interference v separovaném signálu. V případě kdy  $\beta = 0,5$  dojde k potlačení interferencí a hodnota PESQ se zvýší. V případě kdy  $\beta = 0$  dochází již k degradaci hudebním šumem a hodnota PESQ se opět sníží. Výjimkou jsou systémy M-LCMV a M-Max SNR v UDC uspořádání, kde jsou i v případě  $\beta = 0,5$  interference výraznější, a proto je možné dalším snížením  $\beta$  dosáhnout lepších výsledků. Při nastavení  $\beta = 0$  však dosahují všechny systémy využívající metod BF lepších výsledků než MENUET. A to i přes to, že se jedná o systémy s časově-frekvenčním binárním maskováním, jako je tomu v případě systému MENUET.

Na základě výsledků z tab. 8.1 byly vybrány pro subjektivní testy systémy MENUET, FMMENUET (1,5), M-Max SNR, BA1 a BA2. Aby byla zachována linearita BSS systému nebude využito jednokanálové zvýraznění z kapitoly 7. Výjimkou bude pouze systém M-Max SNR v UDC uspořádání, kde nastavíme  $\beta = 0$ .

Pro subjektivní testy byl zvolen CCR test [21] při kterém se porovnává testovaný signál s referenčním signálem. V počátečních fázích subjektivních testů byly prováděny experimenty také s MOS testem [32], kde probíhá hodnocení pouze na základě testovaného signálu. Posluchači však nebyli schopni bez referenčního signálu zachytit rozdíly mezi testovanými signály, a proto bylo od tohoto testu upuštěno.

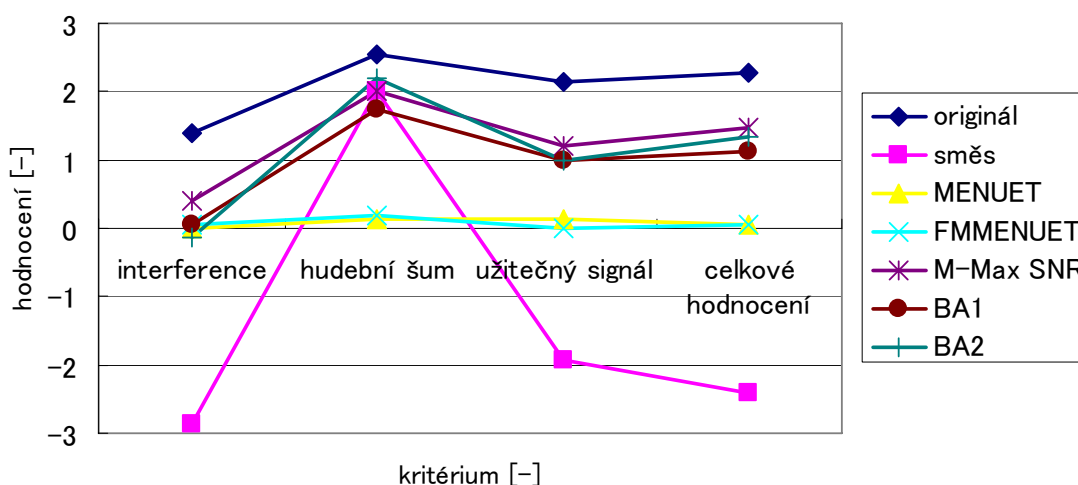
CCR testy hodnotí testovaný signál jako celek jednou známkou, a to na základě porovnání s referenčním signálem.

známka	slovní vyjádření
3	o mnoho lepší
2	lepší
1	o trochu lepší
0	stejně
-1	o trochu horší
-2	horší
-3	o mnoho horší

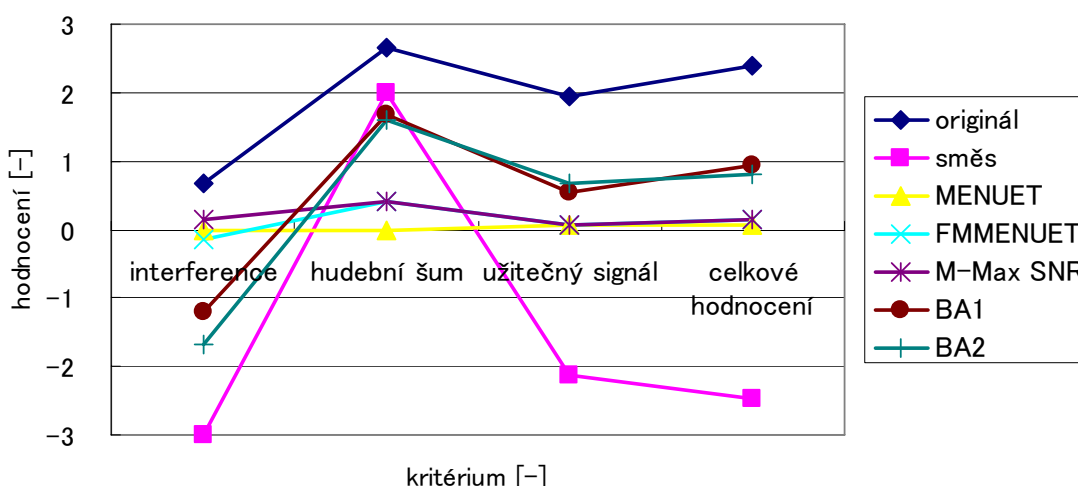
V rámci této disertační práce rozšíříme CCR testy oproti doporučení ITU-T P.800 [21] podle doporučení ITU-T P.835 k provádění MOS testů [32], kde se kromě celkového hodnocení přidává hodnocení užitečného signálu a rušení. Některé z představených BSS systémů však vnášejí do separovaného signálů kromě interferencí také hudební šum. Proto bylo rozděleno rušení na dvě testované kategorie, a to interference a hudební šum. Celkově se tedy separovaný signál bude hodnotit pomocí čtyř kritérií: interference, hudební šum, užitečný signál a celkové hodnocení.

Test se skládal z 26 dvojic srovnávaných souborů a může být nalezen na příloženém CD. Test byl vyhodnocen 15 posluchači pracujícími v oboru zpracování signálů. Referenční signál tvořil separovaný signál získaný systémy MENUET a BA2 v DC i UDC uspořádání. Kromě dříve vybraných BSS systémů byl do testu zahrnut originální signál a směs signálů, jež jsou pozorovány na senzorech.

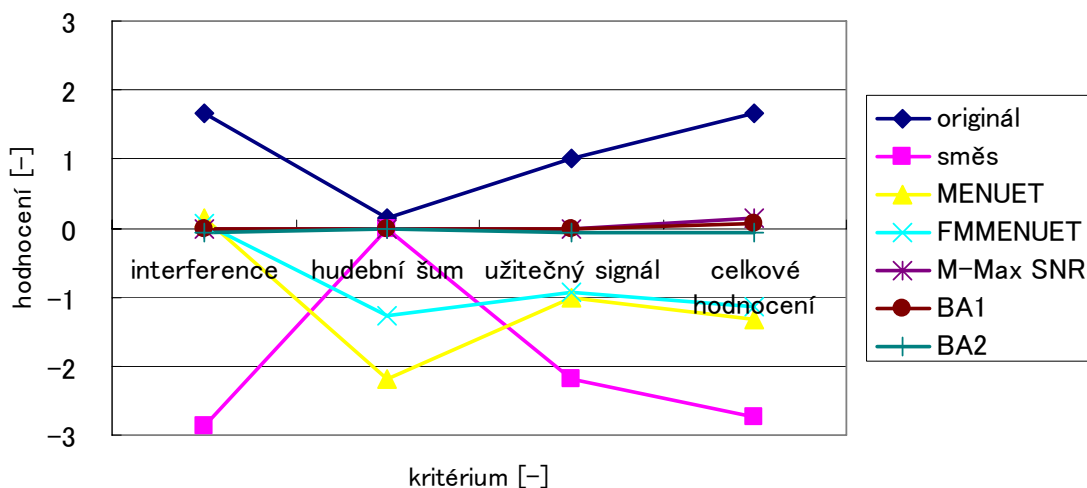
Testované signály byly dlouhé 6 sekund. Z důvodu vyhodnocování rušení byl na začátku i na konci užitečný signál po 1 sekundu neaktivní [32]. Testované signály byly získány během experimentu, a to za totožných podmínek jako při vyhodnocování PESQ. Rozdíl byl pouze v počtu signálů. Z  $N$  zdrojových signálů byl vyhodnocován pouze jeden, a sice mužská resp. ženská promluva v DC resp. UDC uspořádání. Výsledky subjektivního testu jsou zobrazeny na obr. 8.1 až obr. 8.4. Ve formě tabulkového přehledu jsou výsledky testů uvedeny v příloze.



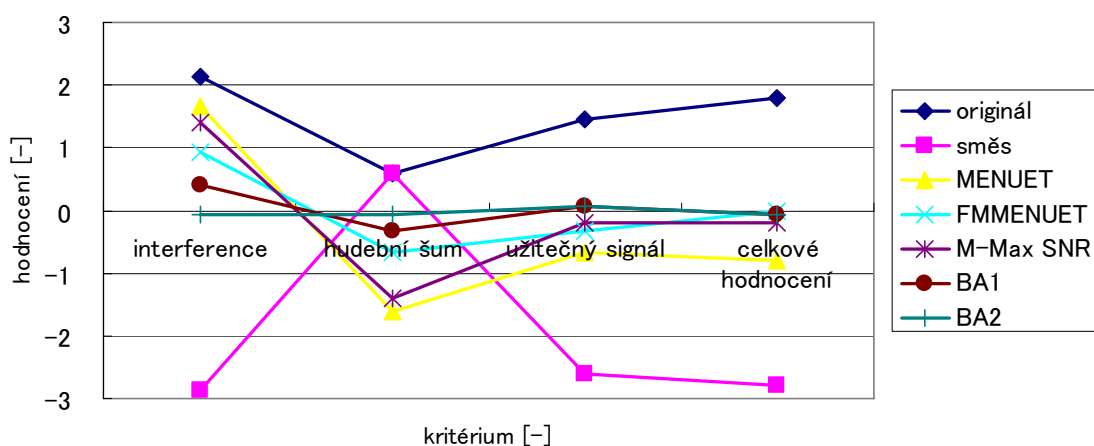
Obr. 8.1. Výsledky subjektivního testu. DC uspořádání. Referenční signál – separovaný signál systémem MENUET.



Obr. 8.2. Výsledky subjektivního testu. UDC uspořádání. Referenční signál – separovaný signál systémem MENUET.



Obr. 8.3. Výsledky subjektivního testu. DC uspořádání. Referenční signál – separovaný signál systémem BA2.



Obr. 8.4. Výsledky subjektivního testu. UDC uspořádání. Referenční signál – separovaný signál systémem BA2.

Originální signál dosáhl nejlepších výsledků jak při porovnání se systémem MENUET, tak s BA2. Je zřejmé, že až na interference dosahuje originální signál při srovnání s BA2 nižších výsledků než při porovnání s systémem MENUET. Směs signálů byla naproti tomu hodnocena nejhůře. Dobrých výsledků dosahuje pouze pro hodnotící kritérium hudební šum. Systém MENUET a BA2 při porovnání s sebou samým zaznamenává výsledek blízký nule pro všechna hodnotící kritéria, což vypovídá o věrohodnosti výsledků. FMMENUET zaznamenal vyšší interference a nižší hudební šum než MENUET. Ostatní kritéria byla hodnocena přibližně stejně. Tento výsledek odpovídá charakteristice obou systémů.

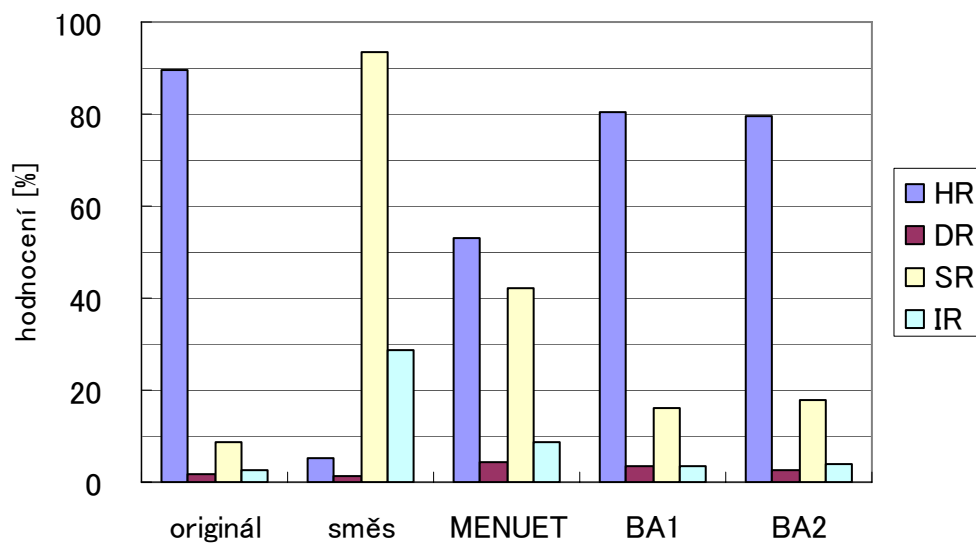
Systémy kombinující tvarovače a MMENUET dosahují v DC uspořádání přibližně stejná hodnocení, a to jak v porovnání se systémem MENUET tak i s BA2. Podobných výsledků jako MENUET dosahují pouze v kritériu interference. V ostatních kritériích jsou hodnoceny lépe. V UDC uspořádání je u M-Max SNR použito jednokanálové binární maskování. I přes to je hudební šum u tohoto systému hodnocen lépe než u systému MENUET. V ostatních kritériích se systémy přibližně shodují.

Nejlepších výsledků z separovaných signálů dosahuje BA2 a BA1. V UDC uspořádání BA1 vykazuje méně interferencí. Posluchačům se zdá být v BA1 více hudebního šumu než v BA2. Oba dva systémy jsou však z hlediska hudebního šumu stejné. Tento výsledek však může být způsoben silnějším potlačením interferencí v případě BA1, kdy zbylé interference připomínají hudební šum.

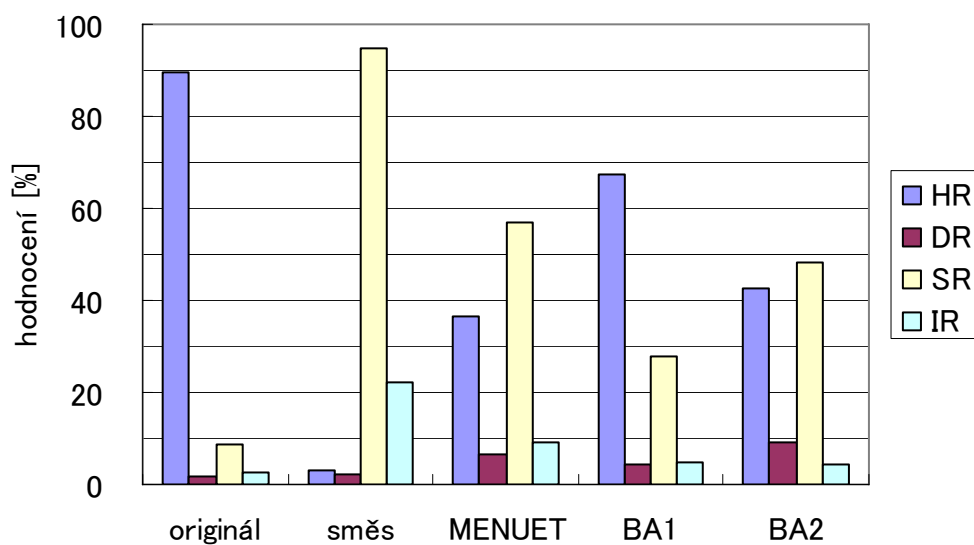
Posledním testem, kterému byly navržené BSS systémy podrobeny, je objektivní vyhodnocení pomocí automatického rozpoznávače řeči (ASR). Objektivní hodnocení kvality řeči pomocí ASR bylo prezentováno např. v [38], [40]. K testu využijeme ASR systém vyvinutý na Technické univerzitě v Liberci [23]. Jedná se o spojitý ASR se slovníkem čítajícím přibližně 300 tisíc slov. Využití nalézá v diktování a automatickém přepisu spojitě řeči.

K testu byly vybrány jako referenční signály originální signál a směs. Z BSS systému pak výstupní signál systému MENUET, BA1 a BA2. Testovací korpus měl celkem 1863 slov, z čehož 14 slov, jak se po provedení testu ukázalo, nebylo ve slovníku ASR systému. ASR test je součástí přiloženého CD. Testované signály byly získány během experimentu, a to za totožných podmínek jako při vyhodnocování PESQ. Rozdíl byl pouze v použití stejné mužské promluvy jak pro DC tak UDC uspořádání. Testované signály byly převzorkovány z 8 kHz na 16 kHz, neboť ASR systém pracuje standardně s  $f_s = 16$  kHz. Výsledky testů pro DC resp. UDC uspořádání jsou uvedeny na obr. 8.5 resp. obr. 8.6, kde HR představuje počet správně rozpoznaných slov, DR počet vynechaných slov, SR počet zaměněných slov a IR počet nesprávně vložených slov. Hodnoty HR, DR, SR a IR jsou vyjádřeny v procentech z celkového počtu slov. Výsledky testů ve formě tabulkového přehledu jsou uvedeny v příloze.





Obr. 8.5. Výsledky ASR testu. DC uspořádání.



Obr. 8.6. Výsledky ASR testu. UDC uspořádání.

V DC i UDC uspořádání byl použit stejný mluvčí. Pozice mluvčího vůči senzorovému poli byla také stejná. Výsledky ASR testu originálního signálu jsou tedy totožné pro DC i UDC uspořádání. Směs signálů dosahuje velmi špatného HR hodnocení. Drtivá většina slov je zaměněna. Vysokého hodnocení dosahuje také hodnota IR. Separace systémem MENUET podstatně zlepšuje HR, přesto se dostáváme v DC uspořádání pouze na hodnotu přesahující 50%, což je oproti originálnímu signálu

značný pokles. V UDC uspořádání nedosahuje HR ani 40%. Vysokých hodnot dosahuje opět hodnota SR. IR zaznamenává přibližně 10%.

BA1 a BA2 dosahují výrazného nárůstu HR a poklesu SR v DC uspořádání. Získané hodnoty obou systémů jsou prakticky srovnatelné. V UDC uspořádání však BA1 výrazně předčí BA2. Pokud porovnáme tento výsledek se subjektivními testy a PESQ, zjistíme, že ASR testy tyto dva systémy výrazněji odlišují. Tento výsledek je zřejmě způsoben vyšším potlačením interferujících signálu systémem BA1 v časově-frekvenčních složkách, kde  $M_k(f, \tau) = 0$ .

## 9 Závěr

Tato disertační práce se zaměřovala na slepou separaci zvukových signálů. Cílem práce bylo navrhnout systém vícekanálové slepé separace zvukových signálů, jež může být použit i v případě nedourčeného uspořádání a zároveň nedegraduje separované signály hudebním šumem. Systém měl být navržen pro prostředí se slabými reverberacemi a dosahovat vysokého potlačení interferujících signálů při nízkém zkreslení užitečného signálu.

Disertační práce začíná uvedením do problému slepé separace signálů v kapitole 1 a představením současných separačních metod v kapitole 2. Po stanovení cílů práce v kapitole 3 začíná vlastní návrh separačního systému v kapitole 4.

Návrh nového separačního systému vychází ze systému MENUET, jenž je založen na časově-frekvenčním binárním maskování. Systém MENUET umožňuje slepou separaci signálů i v nedourčeném uspořádání. Separované signály jsou však degradovány hudebním šumem. Pro potlačení hudebního šumu byla v kapitole 4.2 navržena separační maska s více stupni příslušnosti. Na základě této nové masky bylo možné navrhnout separační systém s nastavitelnou směrovostí a systém omezující vkládání nulových hodnot do spektra separovaného signálu.

Zavedením separační masky s více stupni příslušnosti však není možné hudební šum zcela odstranit, neboť stále dochází k nelineárním změnám ve spektru signálu, které mohou v menší či větší míře degradovat separované signály hudebním šumem.

Závěry z kapitoly 4 vedly k návrhu lineárního separačního systému v kapitole 5, kde byl představen separační systém kombinující systém MENUET a tvarovače přijímací charakteristiky. Aby byla tato kombinace možná, musely být provedeny v systému MENUET úpravy zajišťující slepý odhad parametrů nutných pro návrh tvarovačů přijímací charakteristiky.

Separace představeným systémem může být provedena dokonce i v nedourčeném uspořádání. Separované signály neobsahují žádný hudební šum, neboť separační proces je lineární. V případě nedourčené separace však dochází k zvýšenému pronikání interferencí do separovaného signálu.

Kapitola 6 se zabývá zachováním linearity separačního procesu a zamezením zvýšenému pronikání interferujících signálů do separovaného signálu v nedourčeném uspořádání. Nejprve byl v kapitole 6.1 představen separační systém převádějící nedourčené uspořádání na určené uspořádání. Tento systém dosahuje vyššího potlačení interferencí na druhou stranu, ale také více zkresluje užitečný signál. Při převodu na určené uspořádání dochází také k nelineárním změnám ve spektru.

V kapitole 6.2 bylo představeno pole tvarovačů přijímací charakteristiky vycházející ze separačního systému z kapitoly 6.1. Pole tvarovačů přijímací

charakteristiky nevnaší do separovaného signálu hudební šum, a to dokonce ani v nedourčeném uspořádání. Zároveň je dosaženo vysokého potlačení interferencí a nižšího zkreslení užitečného signálu než u dříve představených separačních systémů.

U systémů představených v kapitole 5 a 6 je možné dále potlačit interference jednokanálovou technikou založenou na využití separační masky navržené systémem MENUET během separačního procesu. Tato technika je popsána v kapitole 7. Představené jednokanálové zvýraznění signálu však představuje nelineární změny ve spektru, které mohou způsobit zanesení hudební šumu do separovaného signálu. Jedná se však o efektivní techniku s minimální výpočetní náročností, která umožňuje regulovat množství interferencí pronikajících do separovaného signálu.

Nedílnou součástí disertační práce je také srovnání separačních systémů pomocí objektivních metod SIR, SDR, PESQ a ASR testu, ale i subjektivní metody CCR. Výsledky těchto experimentů jsou shrnuty v kapitole 8. Subjektivní testy provedené v této disertační práci poskytují informaci o kvalitě separačního systému z hlediska hudebního šumu, interferencí, zkreslení užitečného signálu a o celkovém dojmu posluchačů. Je třeba však upozornit, že subjektivních testů se zúčastnilo pouze 15 posluchačů. Aby byly výsledky prokazatelné, bylo by třeba použít více posluchačů a poslechových ukázek. Z porovnání výsledků objektivních a subjektivních testů lze však usuzovat, že subjektivní testy jsou vypovídající, neboť přibližně odpovídají výsledkům objektivních testů.

Výsledky testů ukazují, že separační systémy založené na kaskádním spojení systému MENUET a pole tvarovačů přijímací charakteristiky dosahují nejlepších výsledků jak v objektivních tak i v subjektivních testech.

V rámci této disertační práce vznikly nové systémy určené k nedourčené separaci zvukových signálů. Výsledky výzkumu v této oblasti byly prezentovány na mezinárodních konferencích a ve vědeckých časopisech. Kaskádní spojení systému MENUET a tvarovačů přijímací charakteristiky bylo patentováno firmou NTT Corporation. Za hlavní přínos této disertační práce je možné považovat navržení nového separačního systému pro nedourčenou slepou separaci signálů, který nedegraduje separovaný signál hudebním šumem, vykazuje nízké zkreslení užitečného signálu a zároveň účinně potlačuje interferující signály.

V budoucím výzkumu, navazujícím na tuto disertační práci, by bylo vhodné se zaměřit na zvýšení odolnosti separačních systémů založených na časově-frekvenčním maskování, a tím i celého separačního celku, oproti reverberacím. Dalším směrem, jímž by bylo vhodné ve vývoji separačních systémů pokračovat je automatická detekce počtu mluvčích.

# Literatura

## Seznam použité literatury

- [1] ARAKI, S., SAWADA, H., MAKINO, S. Blind Speech Separation in a Meeting Situation with Maximum SNR Beamformers. In *ICASSP2007*. Volume 1, 2007, pp. 41–44. ISBN 1–4244–0728–1.
- [2] ARAKI, S., SAWADA, H., MUKAI, R., MAKINO, S. A Novel Blind Source Separation Method with Observation Vector Clustering. In *IWAENC2005*, 2005, pp. 117–120. ISBN 90–386–2311–9.
- [3] ARAKI, S., SAWADA, H., MUKAI, R., MAKINO, S. DOA Estimation for Multiple Sparse Sources with Normalized Observation Vector Clustering. In *ICASSP2006*. Volume 5, 2006, pp. 33–36. ISBN 1–4244–0469–X.
- [4] ARAKI, S. Convolutional Blind Speech Separation with Independent Component Analysis and Sparse Component Analysis. *Disertační práce*. Hokkaido university, 2007.
- [5] BARTSCH, H.J. *Matematické vzorce*. Mladá fronta, 1996, 832 pp. ISBN 80–204–0607–7.
- [6] BELL, A.J., SEJNOWSKI, T.J. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. In *Neural Computation*. Volume 7, 1995, pp. 1129–1159. ISSN: 0899–7667.
- [7] BOTFILL, P., ZIBULEVSKY, M. Underdetermined Blind Source Separation Using Sparse Representations. In *Signal processing*. Volume 18, issue 11, 2001, pp. 2353–2362.
- [8] CARDOSO, J.F., SOULOUMIAC, A. Blind Beamforming for non-Gaussian Signals. In *IEEE Proceedings-F*. Volume 140, issue 6, pp. 362–370, 1993.
- [9] DUDA, R.O., HART P.E., STORK, D.G. *Pattern Classification*. Wiley-Interscience, 2000, pp. 654. ISBN 047–10–566–93.
- [10] GORRIZ, J. M., RAMÍREZ, J., LANG E. W., PUNTONET, C. G. Hard C-means Clustering for Voice Activity Detection. In *Speech Communication*. Volume 48, issue 12, 2006, pp 1638–1649. ISSN: 0167–6393.
- [11] HERMANISKY, H., MORGAN, N. RASTA Processing of Speech, In *Speech and Audio Processing*. Volume 2, issue 4, 1994, pp. 578–589. ISSN: 1063–6676.
- [12] HU, Y., LOIZOU, P. C. Subjective Comparison and Evaluation of Speech Enhancement Algorithms. In *Speech Communication*. Volume 49, issue 7–8, pp. 588–601, 2007. ISSN: 0167–6393.

- [13] HYVARINEN A. Fast and Robust Fixed-Point Algorithm for Independent Component Analysis. In *IEEE Trans. Neural Networks*. Volume 10, issue 3, pp. 626–634, 1999.
- [14] HYVARINEN A., OJA E. Independent Component Analysis: Algorithms and Applications. In *Neural Networks*. Volume 13, issues 4–5, 2000, pp. 411–430.
- [15] JAN, J. *Číslicová filtrace, analýza a restaurace signálů*. Vutium, 2002, ISBN 80-214-1558-4.
- [16] JOHNSON, D., DUDGEON, D. *Array Signal Processing: Concepts and Techniques*. Simon & Schuster, 1992, pp. 533. ISBN 013–0485–136.
- [17] KOLDOVSKÝ, Z. TICHAVSKÝ, P. Time-domain Blind Audio Source Separation Using Advanced Component Clustering and Reconstruction. In *The Joint Workshop on Hands-free Speech Communication and Microphone Arrays 2008*, Italy, 2008.
- [18] LI, Y., CICHOCKI, A., AMARI, S. Sparse Component Analysis for Blind Source Separation with less Sensors than Sources. In *ICA2003*, Japan, 2003, pp. 89–94.
- [19] MAKINO, S., LEE, T., SAWADA, H. *Blind Speech Separation*. Springer, 2007. ISBN 978–1–4020–6478–4.
- [20] MAKINO, S., SAWADA, H., MUKAI, R., ARAKI S. Blind Source Separation of Convolutional Mixtures of Speech in Frequency Domain. In *IEICE Trans. Fundamentals*. Vol. E88-A, issue. 7, pp. 1640–1655, 2005. ISSN: 0916–8508.
- [21] *Methods for subjective determination of transmission quality*. ITU-T Recommendation P.800. Dostupné z: <http://www.itu.int/rec/T-REC-p>, [cit. 04/2008].
- [22] MURAKAMI, T., HOYA, T., ISHIDA, Y. Speech Enhancement by Spectral Subtraction Based on Subspace Decomposition. In *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*. Volume E88-A, issue 3, 2005, pp. 690–701. ISSN: 0916–8508.
- [23] NOUZA, J., ŽDÁNSKÝ, J., ČERVA, P., KOLORENČ, J. A System for Information Retrieval from Large Records of Broadcast Programs. In *Text, Speech and Dialogue. Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin (2006) pp. 401–408.
- [24] OLSHAUSEN B.A., FIELD D.J. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? In *Vision Research*. Volume 37, issue 23,

1997, pp. 3311–3325.

- [25] *Perceptual evaluation of speech quality (PESQ)*. ITU-T Recommendation P.862. Dostupné z: <http://www.itu.int/rec/T-REC-p>, [cit. 04/2008].
- [26] PLŠEK, M., KÉPESI, M. One-Channel Speech Separation by the Spectrogram Mapping Method. In *Research in Telecommunication Technology RTT 2001*. VUT FEI Brno, 2001, pp. 275–280. ISBN 80–214–1938–5.
- [27] RICKARD, S., BALAN, R., ROSCA, J. Real-Time Time-Frequency Based Blind Source Separation. In *ICA2001*. 2001, pp. 651–656.
- [28] RICKARD, S., MELIA, T., FEARON, C., DESPRIT – Histogram Based Blind Source Separation of More Sources than Sensors Using Subspace Methods. In *Applications of Signal Processing to Audio and Acoustics*. 2005, pp. 5–8. ISBN 0–7803–9154–3.
- [29] SAWADA, H., ARAKI S., MUKAI, R., MAKINO, S. Blind Extraction of a Dominant Source from Mixtures of Many Sources Using ICA and Time-Frequency Masking. In *IEEE Trans. Audio, Speech and Language Processing*. Volume 14, issue 6, 2006, pp. 2165–2173.
- [30] SMÉKAL, Z., SYSEL, P. *Signálové procesory*. Sdělovací technika, 2006, pp. 283. ISBN 80–86645–08–8.
- [31] STONE, J. V. *Independent Component Analysis: A Tutorial Introduction*. MIT Press, 2004, 211 pp. ISBN 0262693151.
- [32] *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*. ITU-T Recommendation P.835. Dostupné z: <http://www.itu.int/rec/T-REC-p>, [cit. 04/2008].
- [33] VAN TREES, H. L. *Optimum Array Processing*. Wiley-Interscience, 2002, pp. 1456. ISBN 047–109–3904.
- [34] VAN VEEN, B.D., BUCKLEY K.M. Beamforming: a Versatile Approach to Spatial Filtering. In *ASSP Magazine*. Volume 5, issue 2, 1988, pp. 4–24.
- [35] VIELVA, L., PEREIRO, Y., ERDOGMUS, D., PRINCIPE, J.C. Inversion Techniques for Underdetermined BSS in an Arbitrary Number of Dimensions. In *ICA2003*. Japan, 2003, pp. 131–136. ISBN 4–9901–5311–1.
- [36] VIELVA, L., ERDOGMUS, D., PRINCIPE, J.C. Underdetermined Blind Source Separation Using A Probabilistic Source Sparsity Model. In *ICA 2001*. USA, 2001, pp. 675–679.
- [37] VÍCH, R., SMÉKAL, Z. *Číslíkové filtry*. Academia, Praha 2000. ISBN 80–200–0761–X.

- [38] VÍCH, R., NOUZA, J., VONDRA, M. Automatic Speech Recognition Used for Evaluation of Text-to-Speech Systems. *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*. LCNS 5042, Springer, 2008, pp. 136–148.
- [39] VONDRA, M., VICH, R. Adaptive Comb Filtering in Speech Enhancement by Spectral Subtraction. In *Elektronische Sprachsignalverarbeitung*. Dresden: TUD press, 2007. pp. 130–136. ISBN 978–3–940046–40–6.
- [40] VONDRA, M.; VÍCH, R. Can ASR be Used for Evaluating Speech Quality?. In *17th Czech-German Workshop Speech Processing*. Praha: UFE AVČR, 2007. pp. 115–121. ISBN 978–80–86269–00–9.
- [41] VONDRA, M. Speech Enhancement Using Cepstral Vocoder. In *17th Czech-German Workshop Speech Processing*. Praha: UFE AV ČR, 2007. pp. 110–114. ISBN 978–80–86269–00–9.
- [42] WINTER, S., KELLERMANN, W., SAWADA, H., MAKINO, S. MAP-Based Underdetermined Blind Source Separation of Convolutional Mixtures by Hierarchical Clustering and L1-norm Minimization. In *EURASIP Journal on Advances in Signal Processing*. Volume 2007, issue 1, 2007, pp. 12. ISSN: 1110–8657.
- [43] YILMAZ, O., RICKARD, S. Blind separation of speech mixtures via time-frequency masking. In *IEEE Transactions on Signal Processing*. Volume 52, issue 7, 2004, pp. 1830–1847.

### **Seznam prací autora vztahujících se k tématu disertační práce**

- [44] ARAKI, S., SAWADA, H., MAKINO, S. ČERMÁK, J. *Blind Signal Extraction Device, Method, Program, and Recording Medium*. Patent application number 2006-231648.
- [45] ČERMÁK, J. Blind Source Separation Using Fuzzy Masking. In *17th Czech – German Workshop*. Praha, UFE AVCR, 2007. pp. 83–87. ISBN 978–80–86269–00–9.
- [46] ČERMÁK, J., ARAKI, S., SAWADA, H., MAKINO, S. Blind Source Separation by Combining Beamformers and a Time-Frequency Binary Mask. In *IWAENC2006*. Paris: Télécom Paris, 2006, pp. 2–5. ISBN 1–4244–0469–X.
- [47] ČERMÁK, J., ARAKI, S., SAWADA, H., MAKINO, S. Blind Source Separation Based on a Beamformer Array and Time-Frequency Binary



- Masking. In *ICASSP2007*. Volume 1, 2007, pp. 145–148. ISBN 1–4244–0728–1.
- [48] ČERMÁK, J., ARAKI, S., SAWADA, H., MAKINO, S. Musical Noise Reduction in Time-Frequency-Binary-Masking-Based Blind Source Separation Systems. In *16th Czech-German Workshop*. Praha, URE AVCR, 2006, pp. 45–50. ISBN 8626–9159.
- [49] ČERMÁK, J., BODEČEK, K. Fuzzy Logic Utilization in Text-to-Speech Systems. In *Telecommunications and Signal Processing 2005*. Brno, VUT, 2005, pp. 183–186. ISBN 80–214–2972–0.
- [50] ČERMÁK, J., SMÉKAL, Z. Underdetermined Blind Source Separation using Linear Separation System. *Multimodal Signals: Cognitive and Algorithmic Issues*. LCNS, Springer. In print.
- [51] ČERMÁK, J., SMÉKAL, Z., VRÁNA, J. Optimalizace FIR filtru na signálovém procesoru typu VLIW Motorola MSC8101. In *Elektrorevue – Internetový časopis* (<http://www.elektrorevue.cz>). Roč. 2004, č. 32, pp. 1–10. ISSN: 1213–1539.
- [52] ČERMÁK, J., STEJSKAL, V. Using Fuzzy System for Prosody Modelling. In *15th Czech-German Workshop*. Praha, 2005, pp. 84–90. ISBN 3938–86317–X.
- [53] EKSLER, V., ČERMÁK, J., ZEŽULA, R. Microphone Arrays for Multiple Sources Localization. In *Telecommunications and Signal Processing 2004*. Brno, VUT, 2004, pp. 137–140. ISBN 80–214–2684–5.
- [54] SMÉKAL, Z., SYSEL, P., ČERMÁK, J. Optimization of Parallel Processing in VLIW Digital Signal Processors. In *Proceedings of the 10th International Workshop on Systems, Signals and Image Processing*. Praha, 2003, pp. 280–283. ISBN 80–86645–05–3.
- [55] VRÁNA, J., ČERMÁK, J., ZEŽULA, R. Výpočet vlnkové transformace pomocí algoritmu "lifting". *Elektrorevue – Internetový časopis* (<http://www.elektrorevue.cz>), 2004, č. 34, pp. 1–11. ISSN: 1213–1539.

## Seznam použitých symbolů

$\beta$	koeficient zvýraznění
$\delta_k(f, \tau)$	je Euklidovská vzdálenost mezi $\bar{\mathbf{X}}(f, \tau)$ a $\mathbf{c}_k$
$\theta_T$	bod přechodu masky
$\theta_k$	množina indexů interferujících signálů $k$ -tého zdrojového signálu
$\lambda_{\min}$	minimální délka vlny pozorovaného signálu
$\mu$	průměrná hodnota SIR a SDR souboru
$\xi_k(f, \tau)$	maska jednokanálové metody potlačení interferencí v separovaném signálu
$\sigma$	směrodatná odchylka SIR a SDR souboru
$\tau$	označení rámce
$\mathfrak{I}$	účelová funkce
$*$	konjugace
$(f, \tau)$	časově-frekvenční složky
$\mathbf{A}_k(f)$	prostorový filtr kompenzující vážení u Max SNR tvarovače
$b$	koeficient prolínání
$c$	rychlost zvuku
$c_{kj}$	střed shluků $k$ -tého zdrojového signálu na senzoru $j$
$\mathbf{c}_k$	vektor středů shluků $k$ -tého zdrojového signálu
$C_k$	shluk $k$ -tého zdrojového signálu
$d$	index označující řádku v BA
$D$	počet řádek BA
$d_j$	vzdálenost od zdroje signálu k senzoru $j$
$\mathbf{D}(f)$	diagonální matice ovlivňující energii separovaných signálů
$d_{\max}$	maximální vzdálenost mezi senzory sensorového pole
$E[]$	střední hodnota
$f$	diskrétní frekvence
$f_{\max}$	nejvyšší kmitočet spektra
$f_s$	vzorkovací kmitočet
$g$	strmost přechodu

$h_{kj}$	váha prostředí od zdroje $k$ k senzoru $j$
$\mathbf{h}$	směsná matice
$\mathbf{H}$	konjugace a transpozice
$h_{kj}(l)$	impulsní odezva od zdroje $k$ k senzoru $j$
$H_{kj}(f)$	impulsní odezva od zdroje $k$ k senzoru $j$ ve frekvenční oblasti
$\mathbf{H}_k(f)$	směsný vektor od zdrojového signálu $k$ ke všem sensorům ve frekvenční oblasti
$\mathbf{H}(f)$	směsná matice ve frekvenční oblasti
$\hat{H}_{kj}(f)$	odhad impulsní odezvy od zdroje $k$ k senzoru $j$ ve frekvenční oblasti
$\hat{\mathbf{H}}_k(f)$	odhad směsného vektoru zdrojového signálu $k$ ve frekvenční oblasti
$H(y_k)$	marginální entropie
$H(\mathbf{y})$	sdužená entropie
$I$	řád FIR filtru
$I(\mathbf{y})$	vzájemná informace
$\mathbf{I}$	jednotková matice
$j$	imaginární jednotka
$j$	index senzoru
$J$	referenční senzor
$k$	index zdrojových signálů, index sloupce v BA
$L$	délka impulsní odezvy
$M$	počet sensorů
$M_k(f, \tau)$	maska separačního systému
$N$	počet zdrojových signálů
$\hat{N}_{kj}(f, \tau)$	$j$ -tá odhadnutá směs interferujících signálů zdrojového signálu $k$
$\hat{\mathbf{N}}_k(f, \tau)$	vektor odhadů směsí interferujících signálů zdrojového signálu $k$
$\hat{\mathbf{N}}_{kd}(f, \tau)$	vektor odhadu směsí interferujících signálů $k$ -tého sloupce a $d$ -tého řádku BA
$P$	libovolný senzor
$p(y_k(q))$	model hustoty rozdělení pravděpodobnosti zdrojového signálu
$\mathbf{P}(f)$	permutační matice

$q$	iterační krok
$\mathbf{R}(f)$	korelační matice pozorovaného signálu
$\hat{\mathbf{R}}_{kd}(f)$	odhad korelační matice interferencí pro $d$ -tý řádek a $k$ -tý sloupec BA
$\mathbf{R}_T^k(f)$	korelační matice zdrojového signálu
$\mathbf{R}_N^k(f)$	korelační matice interferencí
$\hat{\mathbf{R}}_N^k(f)$	odhad korelační matice interferencí
$\hat{\mathbf{R}}_T^k(f)$	odhad korelační matice zdrojového signálu
$s_k(t)$	$k$ -tý zdrojový signál
$\mathbf{s}(t)$	vektor zdrojových signálů
$S_k(f, \tau)$	zdrojový signál v časově-frekvenční oblasti
$\mathbf{S}(f, \tau)$	vektor zdrojových signálů v časově-frekvenční oblasti
$\text{SDR}^k$	poměr $k$ -tého užitečného signálu k zkreslení zavedené BSS systémem do $k$ -tého užitečného signálu
$\text{SIR}^k$	potlačení interferencí v $k$ -tém signálu BSS systémem
$t$	diskrétní čas
$T$	$T$ -bodová STFT
$^T$	transpozice
$T_R$	doba dozvuku
$w(t)$	váhová posloupnost k zmírnění Gibbsonova jevu
$\mathbf{W}(f)$	separační matice ve frekvenční oblasti
$\mathbf{W}_k(f)$	vektor FIR filtrů $k$ -tého tvarovače
$W_{kj}(f)$	FIR filtr $k$ -tého tvarovače v $j$ -té větvi
$x_j(t)$	pozorovaný signál na $j$ -tém senzoru
$\mathbf{x}(t)$	vektor pozorovaných signálů
$\mathbf{X}(f, \tau)$	vektor pozorovaných signálů v časově-frekvenční oblasti
$X_{kj}(f, \tau)$	pozorovaný signál na senzoru $j$ od $k$ -tého zdroje v časově frekvenční oblasti
$\bar{\mathbf{X}}(f, \tau)$	normalizovaný vektor v časově-frekvenční oblasti
$\bar{X}_j(f, \tau)$	normalizovaný pozorovaný signál na senzoru $j$
$\tilde{\mathbf{X}}_k(f, \tau)$	vektor vstupních signálů tvarovače
$\tilde{\mathbf{X}}_{kd}(f, \tau)$	vektor vstupních signálů $k$ -tého sloupce a $d$ -tého řádku BA

$y_k(t)$	separovaný signál
$\mathbf{Y}(f, \tau)$	vektor separovaných signálů v časově-frekvenční oblasti
$Y_k(f, \tau)$	$k$ -tý separovaný signál v časově-frekvenční oblasti
$\hat{\mathbf{Y}}_k(f, \tau)$	vektor odhadů $k$ -tého zdrojového signálu $M$ senzorech
$\hat{Y}_{kj}(f, \tau)$	odhad $k$ -tého signálu na senzoru $j$
$\hat{\mathbf{Y}}(f, \tau)$	matice odhadů zdrojových signálů
$\tilde{Y}_{kd}(f, \tau)$	výstupní signál $k$ -tého sloupce a $d$ -tého řádku BA
$Z$	počet interferujících signálů obsažených ve vektoru vstupních signálů $\tilde{\mathbf{X}}_{kd}(f, \tau)$
$\mathbf{z}_k$	množina s $M - 1$ indexy interferujících signálů
$\mathbf{z}_{kd}$	$d$ -tá $Z$ -prvková podmnožina množiny $\theta_k$

## Seznam zkratek

2D	dvojrozměrný prostor
ASR	automatické rozpoznávání řeči (Automatic Speech Recognition)
BA	pole tvarovačů (Beamformer Array)
BF	tvarování přijímací charakteristiky (BeamForming)
BSS	slepá separace signálů (Blind Source Separation)
CA	shluková analýza (Cluster Analysis)
CCR	hodnocení porovnáním (Comparison Category Rating)
DC	určená separace (počet signálů je roven počtu senzorů), (Determined Case)
DOA	úhel dopadu signálu na senzor (Direction Of Arrival)
DR	počet vynechaných slov (Delece Rate)
DUET	separační systém na bázi TFBM využívající pouze dvou senzorů (Degenerate Unmixing Estimation Technique)
FastICA	rychlý algoritmus výpočtu ICA v pevné řádové čárce
FIR	číslicový filtr s konečnou impulsní charakteristikou (Finite Impulse Response)
FMMENUET	separační systém MENUET s fuzzy maskou (Fuzzy Mask MENUET)
HR	počet správně rozpoznaných slov (Hit Rate)
ICA	analýza nezávislých komponent (Independent Component Analysis)
IR	počet nesprávně vložených slov (Insertion Rate)
ISTFT	inverzní krátkodobá Fourierova transformace (Inverse STFT)
JADE	algoritmus výpočtu ICA (Joint Approximate Diagonalization of Eigenmatrices)
LCMV	metoda návrhu statisticky optimálního tvarovače nazývaná lineárně podmíněná minimální variance (Linearly Constrained Minimum Variance)
MAP	maximální aposteriorní pravděpodobnost (Maximum A Posteriori)
Max SNR	metoda maximálního odstupe signálu od šumu používaná k návrhu statisticky optimálního tvarovače (Maximal Signal to Noise Ratio)
MENUET	separační systém založený na systému DUET umožňující využívat

	více senzorů s libovolným uspořádáním (Multiple sENzor dUET)
M-LCMV	separační systém kombinující MMENUET a LCMV tvarovač
M-Max SNR	separační systém kombinující MMENUET a Max SNR tvarovač
MMENUET	vícekanálový MENUET (Multi channel MENUET)
MSC	metoda návrhu statisticky optimálního tvarovače nazývaná vícenásobný potlačovač postraních laloků (Multiple Side-lobe Canceller)
PCA	analýza hlavních komponent (Principal Component Analysis)
PESQ	objektivní metoda hodnocení hlasové kvality (Perceptual Evaluation of Speech Quality)
RASTA	jednokanálová metoda zvýrazňování řeči (RelAtive SpecTrAl technique)
SCA	analýza řídkých komponent (Sparse Component Analysis)
SDR	soubor hodnot $\text{SDR}^k$ (Signal to Distortion Ratio)
SIR	soubor hodnot $\text{SIR}^k$ (Signal to Interference Ratio)
SMMENUET	MENUET s měkkou maskou (Soft Mask MENUET)
SR	počet zaměněných slov (Substitution Rate)
STFT	krátkodobá Fourierova transformace (Short Time Fourier Transform)
TFBM	časově-frekvenční binární maskování (Time-Frequency Binary Masking)
UDC	nedourčená separace (počet signálů převyšuje počet senzorů), (UnderDetermined Case)
VAD	detektor řečové aktivity (Voice Activity Detection)

## **Seznam příloh**

Příloha A – Výsledky subjektivního testu.

Příloha B – Výsledky ASR testu.

Příloha C – Obsah přiloženého CD.



## Příloha A – Výsledky subjektivního testu

Tab. P.1. Výsledky pro DC uspořádání, referenční signál – MENUET.

	interference	hudební šum	užitečný signál	celkové hodnocení
originál	1,4	2,5	2,1	2,3
směs	-2,9	2,0	-1,9	-2,4
MENUET	0,0	0,1	0,1	0,1
FMMENUET	0,1	0,2	0,0	0,1
M-Max SNR	0,4	2,0	1,2	1,5
BA1	0,1	1,7	1,0	1,1
BA2	-0,1	2,2	1,0	1,3

Tab. P.2. Výsledky pro UDC uspořádání, referenční signál – MENUET.

	interference	hudební šum	užitečný signál	celkové hodnocení
originál	0,7	2,7	1,9	2,4
směs	-3,0	2,0	-2,1	-2,5
MENUET	0,0	0,0	0,1	0,1
FMMENUET	-0,1	0,4	0,1	0,1
M-Max SNR	0,1	0,4	0,1	0,1
BA1	-1,2	1,7	0,5	0,9
BA2	-1,7	1,6	0,7	0,8

Tab. P.3. Výsledky pro DC uspořádání, referenční signál – BA2.

	interference	hudební šum	užitečný signál	celkové hodnocení
originál	1,7	0,1	1,0	1,7
směs	-2,9	0,0	-2,2	-2,7
MENUET	0,1	-2,2	-1,0	-1,3
FMMENUET	0,1	-1,3	-0,9	-1,1
M-Max SNR	0,0	0,0	0,0	0,1
BA1	0,0	0,0	0,0	0,1
BA2	-0,1	0,0	-0,1	-0,1

Tab. P.4. Výsledky pro UDC uspořádání, referenční signál – BA2.

	interference	hudební šum	užitečný signál	celkové hodnocení
originál	2,1	0,6	1,5	1,8
směs	-2,9	0,6	-2,6	-2,8
MENUET	1,7	-1,6	-0,7	-0,8
FMMENUET	0,9	-0,7	-0,3	0,0
M-Max SNR	1,4	-1,4	-0,2	-0,2
BA1	0,4	-0,3	0,1	-0,1
BA2	-0,1	-0,1	0,1	-0,1

## Příloha B – Výsledky ASR testu

Tab. P.5. Výsledky pro DC uspořádání.

	originál	směs	MENUET	BA1	BA2
HR [%]	89,5	5,1	53,2	80,5	79,6
DR [%]	1,7	1,4	4,5	3,4	2,7
SR [%]	8,8	93,5	42,2	16,1	17,7
IR [%]	2,6	28,9	8,6	3,3	3,9

Tab. P.6. Výsledky pro UDC uspořádání.

	originál	směs	MENUET	BA1	BA2
HR [%]	89,5	3,2	36,6	67,6	42,6
DR [%]	1,7	2,1	6,3	4,5	9,1
SR [%]	8,8	94,7	57,1	28,0	48,3
IR [%]	2,6	22,1	9,2	4,8	4,2

## **Příloha C – Obsah přiloženého CD**

Organizace přiloženého CD je následující:

<b>jméno adresáře</b>	<b>popis</b>
ASR_test	Devět ASR testů. Každý z testů se skládá ze 42 souborů separované promluvy ve formátu wav a z přepisu těchto promluv ve formátu txt.
subjektivni_test	Subjektivní text včetně odpovědného formuláře, ukázky hudebního šumu a interferencí.
text_prace	Disertační práce v pdf formátu.

# Životopis

## Osobní údaje:

Jméno:	Ing. Jan Čermák
Datum narození:	10. 9. 1979
Kontakt:	cermak4@kn.vutbr.cz

## Vzdělání:

2003 – 2008	Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Obor Teleinformatika. Postgraduální doktorské studium.
2003 – 2007	Vysoké učení technické v Brně. Fakulta podnikatelská. Obor Ekonomika a řízení podniku. Magisterské studium.
1998 – 2003	Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Obor Elektronika a sdělovací technika. Magisterské studium.
2001 – 2002	Pforzheim University of Applied Sciences, Německo. Fakulta Elektrotechniky a Informatiky. Čtvrtý ročník magisterského studia.
1994 – 1998	Střední průmyslová škola v Písku. Obor Telekomunikace.

## Praxe:

2005 – 2008	Ústav fotoniky a elektroniky, Akademie věd ČR. Zpracování řečových signálů.
2006	NTT Corporation, NTT Communication Science Laboratories, Japan. Slepá separace signálů.
2004	Hilti A.G., Vývojové centrum, Lichtenštejnsko. Vývoj software a hardware pro mechatronické systémy.

**Účast na projektech:**

2007 – 2010	Účast na projektu COST2102 M3MT OC 2102 „Analýza, modelování a syntéza stylů a emočních stavů v řečovém projevu.“
2005 – 2008	Účast na projektu GAAV „Informační společnost“ 1ET301710509 „Sofistikované potlačovače hluku a poruch při přenosu řečových signálů pro pevné a mobilní sítě nové generace.“
2005 – 2006	Hlavní řešitel projektu FRVŠ 2456/2005/G1 „Modelování prozodie TTS systémů fuzzy logikou.“
2003 – 2005	Účast na projektu COST277 MŠMT OC 277.002 „Nelineární metody zvýrazňování řeči.“
2004 – 2005	Účast na projektu FRVŠ 1637/2004/G1 „Senzorová pole k lokalizaci akustických zdrojů v 3D.“